

DE MONTFORT UNIVERSITY, LEICESTER

**Comparison of methyl eugenol levels and
eugenol O-methyltransferase gene structure
in different *Ocimum* plant species**

Purvi Mali

**Biomolecular Technology Group
Faculty of Health and Life Sciences**

PhD

April 2019

Abstract

Ocimum tenuiflorum L. (*Holy basil* or *Tulsi*), belonging to Lamiaceae family, is an important medicinal plant, with particular religious significance to the Hindu community throughout the world. Tulsi plants are characterised by high levels of essential oils, containing phenylpropanoids, such as eugenol, methyl eugenol (ME), chavicol and estragole (methyl chavicol). Two chemotypes of Tulsi have been distinguished, based on high or low methyl eugenol:eugenol ratios. As methyl eugenol and methyl chavicol are classed as genotoxic carcinogens, it is important to ensure that the levels of these compounds in herbal products fall below the regulatory thresholds.

The levels of methyl eugenol in *O. tenuiflorum* are generally higher than in other *Ocimum* species. The conversion of eugenol to methyl eugenol is catalysed by *eugenol O-methyltransferase (EOMT)* enzyme. *EOMT* gene sequences have been isolated from a range of *Ocimum* species and from different chemotypes of *O. tenuiflorum*. Analyses of *EOMT* genomic and cDNA sequences revealed a 843 bp open reading frame and the presence of a 90-104 bp intron. Alignment of the protein sequences from several *Ocimum* species indicates a number of amino acid substitutions that may be correlated with methyl eugenol content and reflect differences in enzyme activity and substrate specificity. Later, the levels of ME along with other three secondary metabolites - eugenol, chavicol and estragole have been determined by HPLC.

Full length *EOMT* cDNA were prepared from *O. tenuiflorum*, *O. gratissimum* and *O. basilicum* plants and are being used to study the properties of the plant *EOMT* enzyme after expression in bacteria. Along with protein study, the Gateway cloning system was successfully used to prepare four binary vectors [*O. basilicum (EOMT)*, *O. basilicum (CVOMT)*, *O. tenuiflorum* and *O. gratissimum*] for future plant transformation experiments.

Contents

Abstract.....	1
1 Introduction	16
1.1 Medicinal Plants: Global perspective	17
1.2 Medicinal Plants: Indian perspective.....	19
1.2.1 Ayurveda	19
1.2.2 Siddha	22
1.2.3 Unani	23
1.2.4 Folk and Sowa-Rigpa	24
1.3 Global herbal industry and chequered history of herbal products	24
1.3.1 The Use of Medicinal Plants by Migrant People.....	25
1.3.2 Current status of Indian traditional medicinal plants standards.....	27
1.3.3 Regulation of Indian Herbals	28
1.4 Plant identification methods	29
1.4.1 Plant morphology	29
1.4.2 The classification system	33
1.4.3 Chemical analysis	34
1.4.4 Plant DNA barcoding	36
1.4.5 Identifying single loci.....	38
1.5 Modern trends in medicinal plants field.....	40
1.5.1 Combinatorial biosynthetic technology	41
1.6 Plant and plant material used for the study	42
1.6.1 Fresh samples	43
1.6.2 Dry sample.....	43
1.6.3 Commercial products	44
1.6.4 Other plant DNA sources.....	44
1.7 Secondary metabolites.....	45
1.7.1 Eugenol (C ₁₀ H ₁₂ O ₂).....	47
1.7.2 Methyl eugenol (C ₁₁ H ₁₄ O ₂)	48
1.7.3 Chavicol (C ₉ H ₁₀ O).....	48
1.7.4 Methyl chavicol (C ₁₀ H ₁₂ O).....	48
1.8 Lamiaceae family plants	48
1.9 The genus <i>Ocimum</i>.....	50
1.9.1 <i>Ocimum tenuiflorum</i> L.	51
1.9.2 <i>Ocimum basilicum</i>	52
1.9.3 <i>Ocimum gratissimum</i>	53
1.9.4 Other <i>Ocimum</i> species	54
1.10 Phylogenetic inference: concepts and methods.....	54
1.11 Research aims and objectives	55
1.11.1 Explore the occurrence of DNA sequence polymorphisms within the key biosynthetic genes for ME and estragole pathway	57
1.11.2 Determining the effect of genetic differences on enzyme activity and substrate specificity	58
2 General materials and methods.....	59
2.1 Plant Material collection and storage.....	59

2.1.1	<i>Ocimum</i>	59
2.1.2	<i>Arabidopsis</i>	59
2.1.3	Prior procedure on samples before extracting DNA	59
2.2	Seed germination	60
2.3	Genomic DNA extraction	60
2.4	RNA extraction and cDNA synthesis	61
2.5	Polymerase Chain Reaction (PCR)	62
2.5.1	Primer design	62
2.6	Gel Electrophoresis	63
2.7	Isopropanol Clean Up	63
2.8	Purification of PCR products from gels	64
2.9	PCR purification	64
2.10	Bioinformatics tools	64
2.10.1	Gel Purification	65
3	Authentication	66
3.1	Introduction	66
3.1.1	DNA barcoding	67
3.1.2	Advantage of using DNA based identification techniques	68
3.1.3	DNA test introduction into the British Pharmacopoeia	69
3.1.4	Authentication of different <i>Ocimum</i> species	70
3.1.5	High Performance Thin Layer Chromatography (HPTLC)	71
3.1.6	DNA authentication of different <i>Ocimum</i> species	73
3.1.7	Steps required for DNA barcoding	73
3.2	Materials and methods	76
3.2.1	Plant Sample materials	76
3.2.2	DNA samples	77
3.2.3	DNA isolation	80
3.2.4	PCR to test DNA extractions and Plant barcoding	81
3.2.5	Sample sequencing	82
3.3	Result	82
3.3.1	Gel electrophoresis: <i>trnH-psbA</i> region	82
3.3.2	DNA authentication of unknown Tulsi samples	83
3.4	Discussion	89
3.4.1	<i>trnH-psbA</i>	89
3.4.2	Advantage of using DNA based technique	90
3.4.3	In detail investigation	91
3.4.4	Contig description	91
3.4.5	Authentication of unknown samples	92
3.5	Conclusion	93
3.5.1	Further work	94
4	DNA and Amino Acid sequence analysis of <i>O</i>- methyltransferase genes	95
4.1	Introduction	95
4.1.1	Methylation of biological molecules	95
4.1.2	Gene expression	97

4.1.3	Plant Genomics and Bioinformatics	99
4.2	Material and methods.....	102
4.2.1	Primer design	102
4.2.2	DNA samples	104
4.2.3	Use of Bioinformatics tools	105
4.2.4	CLC software.....	106
4.3	Result	108
4.3.1	Primers	108
4.3.2	Nucleotide alignment	110
4.3.3	Visualising Introns	116
4.3.4	AA alignment using nucleotide sequences.....	119
4.3.5	Phylogenetic analysis	134
4.4	Discussion.....	136
4.4.1	Obtaining the consensus sequences	136
4.4.2	Full-length <i>O-methyltransferase</i> gene study	138
4.4.3	Phylogenetic analysis of OMT gene within various <i>Ocimum</i> species	140
4.4.4	The key discoveries of sequence analysis	141
4.4.5	Combination of sequence analysis and phylogenetic study.....	142
4.5	Conclusion	142
4.5.1	Further work.....	143
5	<i>Acquisition and cloning of full length cDNAs representing OMT genes from different Ocimum species.....</i>	144
5.1	Introduction.....	144
5.1.1	<i>O-methyltransferase</i> enzymes (comparisons of <i>EOMT</i> and <i>CVOMT</i> enzymes).....	144
5.1.2	Overview of DNA cloning	145
5.1.3	<i>Escherichia coli</i> (<i>E. coli</i>) as a model system	146
5.1.4	Cloning technologies for protein expression.....	146
5.1.5	The aim of the chapter	149
5.2	Materials and methods	149
5.2.1	<i>pGEM</i> vector cloning	149
5.2.2	Preparing PCR products using G-blocks.....	153
5.2.3	<i>pET</i> vector cloning	154
5.2.4	Sequence analysis.....	157
5.2.5	IPTG Assay	157
5.3	Result	157
5.3.1	Primers covering initial and end parts of the OMT gene region	157
5.3.2	<i>pGEM</i> vector cloning	159
5.3.3	G-blocks preparation for <i>pET</i> vector cloning.....	162
5.3.4	<i>pET</i> colony screening.....	166
5.3.5	SDS-PAGE.....	170
5.4	Discussion.....	171
5.4.1	RNA extraction and cDNA synthesis of <i>O. tenuiflorum</i> plant sample.....	171
5.4.2	Obtaining of OMT sequences for <i>O. basilicum</i> , <i>O. tenuiflorum</i> and <i>O. gratissimum</i>	172
5.4.3	Designing a full-length cDNA sequence representing OMT gene	172
5.4.4	Colony screening sequences	174
5.4.5	Summary of vector cloning results output	175
5.4.6	Protein production study	176
5.5	Conclusion	176

5.5.1	Further work.....	178
6	<i>O-methyltransferase binary vector preparation for the cloning of novel plant transformation constructs.....</i>	179
6.1	Introduction.....	179
6.1.1	Different gene transfer methods in plants.....	179
6.1.2	Plant transformation	179
6.1.3	Plants transformed with homologues of the OMT gene in various plants.....	181
6.1.4	Successful transformation of <i>Ocimum</i> plant	182
6.2	Material and method	182
6.2.1	Gateway pDONR™221 vector (to create entry clone).....	182
6.2.2	Sterilisation process	184
6.2.3	Medium recipe	184
6.3	Result	185
6.3.1	PCR purification	185
6.3.2	Gateway cloning results	185
6.3.3	Seed collection	194
6.4	Discussion.....	194
6.4.1	Initial stage of plant transformation	195
6.4.2	Gateway cloning – different ways of gene assembly	195
6.5	Conclusion	196
6.5.1	Future work	197
7	<i>Chemical analysis.....</i>	198
7.1	Introduction.....	198
7.1.1	Chemical analysis of secondary plant metabolites	198
7.1.2	Chemical properties and threshold levels of target secondary metabolites	201
7.1.3	Diode array detection (DAD) coupled to HPLC.....	204
7.2	Materials and method.....	204
7.2.1	Chemicals and reagents	204
7.2.2	Samples analysed	205
7.2.3	Sample preparation.....	205
7.2.4	Preparation of calibration standards.....	205
7.2.5	HPLC-DAD analysis.....	206
7.3	Results.....	206
7.3.1	Calibration and validation of HPLC method	206
7.3.2	Optimisation of extraction	210
7.3.3	Analysis of Tulsi extracts	211
7.4	Discussion.....	212
7.5	Conclusion	213
7.5.1	Further work.....	214
8	<i>General Discussion and future work.....</i>	215
8.1	Summary of the research project.....	215
8.1.1	Authentication of the starting material.....	216
8.1.2	Explore the occurrence of <i>OMT</i> DNA polymorphism	216
8.1.3	<i>OMT</i> gene transformation in <i>E. coli</i>	221
8.1.4	The identification of secondary metabolites through chemical analysis	224

8.2	Future work	225
8.2.1	Protein expression and <i>OMT</i> enzyme assay	225
8.2.2	<i>Agrobacterium</i> transformation	226
8.2.3	<i>Arabidopsis</i> and <i>Ocimum</i> plant transformation	226
8.2.4	Examine the expression of key biosynthetic genes under environmental and plant growth conditions which affect ME levels in varieties of Tulsi	227
	Reference	228
	Appendix	258
	Appendix 1 Full nucleotide alignment with reference sequences	258
	Appendix 2 Different size introns splicing from genomic DNAs	261
	Appendix 3 Full length <i>O. tenuiflorum</i> sequence obtained from chapter 4 nucleotide alignment	271
	Appendix 4 Full length <i>O. gratissimum</i> sequence obtained from chapter 4 nucleotide alignment	271
	Appendix 5 G-blocks.....	272
	Appendix 6 <i>Ocimum (EOMT)</i> Sequence Accession Numbers from GenBank	274

Content of Figures

Figure 1-1: Different leaf types.....	31
Figure 1-2: Floral structure	32
Figure 1-3: DNA barcoding system for identification purpose	37
Figure 1-4: Representation of <i>rbcL</i> and <i>matK</i> gene structure	39
Figure 1-5: Structure of the ITS region.....	39
Figure 1-6: Chloroplast genome map showing candidate barcode regions	40
Figure 1-7: Secondary metabolites pathway of Holy basil	47
Figure 1-8: <i>Ocimum tenuiflorum</i>	52
Figure 1-9: <i>Ocimum basilicum</i>	53
Figure 1-10: <i>Ocimum gratissimum</i>	54
Figure 1-11: An overview of beginning of the research project	56
Figure 3-1: Essential oil fingerprint of 10 samples of Holy basil.	72
Figure 3-2: Three different <i>Ocimum</i> varieties.....	80
Figure 3-3: <i>trnH-psbA</i> PCR of <i>Ram</i> , <i>Shyam</i> and <i>Vrinda</i> sample.....	83
Figure 3-4: <i>trnH-psbA</i> PCR for Basil DNA extractions.	83
Figure 3-5: Electropherograms of <i>trnH-psbA</i> sequences of three Tulsi types.....	84
Figure 3-6: BLAST search results.....	85
Figure 3-7: Multiple sequence alignment of several <i>O. gratissimum</i> sequences	88
Figure 3-8: Multiple sequence alignment including standard sequence.	89
Figure 3-9: A Section of a Contig Produced from two Electropherograms.....	92
Figure 4-1: Secondary metabolite pathway.....	97
Figure 4-2: Primer design strategy	103
Figure 4-3: The CLC Main workbench.....	106
Figure 4-4: Process of resolving conflicts.....	107
Figure 4-5: Specific primer pair PCR attempt with three <i>Ocimum</i> varieties	109
Figure 4-6: Gel results for two sets of internal primers	110
Figure 4-7: Sequence assembly parameter dialogue box	111
Figure 4-8: Assembly of forward and reverse reads.	112
Figure 4-9: Resolving conflicts	113
Figure 4-10: An initial alignment containing various <i>Ocimum</i> sequences	114
Figure 4-11: A subset of aligned sequences.....	115
Figure 4-12: A final alignment output	116
Figure 4-13: Identifying and removing introns from genomic DNA sequences	117
Figure 4-14: An example of the splice site consensus sequence.	117
Figure 4-15: Representing intron boundaries by aligning gDNA sequences.....	118
Figure 4-16: Three <i>O. gratissimum</i> samples (769, 1138 and 744)	119
Figure 4-17: Standard translation table	120
Figure 4-18: An example of three forward reading frames.....	120
Figure 4-19: A process to translate nucleotide sequence into AA sequence	122
Figure 4-20: Unmodified AA alignment contains frameshift, stop codon and gaps	124
Figure 4-21: Modified AA alignment after removing gaps	124
Figure 4-22: A final version of AA alignment with reference sequences	126
Figure 4-23: An alignment of G37 sequences	127
Figure 4-24: <i>O. tenuiflorum</i> EOMT AA sequence alignment.....	132

Figure 4-25: An AA alignment	134
Figure 4-26: Phylogenetic tree construction	136
Figure 5-1: Molecular cloning of recombinant DNA	145
Figure 5-2: Illustration of IPTG-induced gene expression	149
Figure 5-3: <i>pGEM®-T Easy</i> vector map	150
Figure 5-4: G-block design (suitable for pGEM).....	153
Figure 5-5: G-block design (suitable for pET).....	153
Figure 5-6: <i>pET-28a(+)</i> Vector Map	154
Figure 5-7: The base pair differences.....	158
Figure 5-8: PCR analysis of different <i>Ocimum</i> genomic DNA samples	158
Figure 5-9: RNA extraction samples. (a) c-DNA sample; Lane 4: negative control	159
Figure 5-10: Comparing Genomic DNA, Turbo DNA clean up and cDNA samples...	159
Figure 5-11: PCR reaction confirmation of ligation products.	160
Figure 5-12: pGEM colony screening using M13 primers	161
Figure 5-13: Screenshot of the BLAST search.	161
Figure 5-14: Chromatogram of annotated regions	162
Figure 5-15: Two thermocycler programmes designed for pET <i>EOMT</i> primers.....	164
Figure 5-16: <i>pET_EOMT</i> and <i>pET_CVOMT</i> primers	164
Figure 5-17: Purified PCR samples containing all four G-blocks	165
Figure 5-18: Restriction digest product of all G-blocks and <i>pET</i> vector	165
Figure 5-19: pET vector PCR colony screening	166
Figure 5-20: PCR test confirmation for selected pET colonies using <i>EOMT</i> primers..	167
Figure 5-21: Chromatogram with annotated regions	168
Figure 5-22: Chromatogram with annotated regions	168
Figure 5-23: Chromatogram with annotated regions	169
Figure 5-24: Chromatogram with annotated regions	169
Figure 5-25: SDS-PAGE result 1	170
Figure 5-26: SDS-PAGE result 2.....	171
Figure 6-1: Different phenylpropene production in plants.	181
Figure 6-2: <i>pDONR™221</i> vector map	183
Figure 6-3: PCR-purified product containing <i>Att</i> sites.	185
Figure 6-4: BP reaction colony screening.....	186
Figure 6-5: LR reaction colony screening.....	187
Figure 6-6: LR colony screening using <i>EOMT Att</i> primers.....	188
Figure 6-7: <i>ATT-F</i> primers in the combination with internal reverse (Int-R) primer ...	188
Figure 6-8: Nucleotide sequence of the PCR-amplified OMT region	190
Figure 6-9: Nucleotide sequence of the PCR-amplified OMT region	191
Figure 6-10: Nucleotide sequence of the PCR-amplified OMT region	192
Figure 6-11: Nucleotide sequence of the PCR-amplified OMT region	193
Figure 6-12: Seed germination.....	194
Figure 7-1: Plant phytochemical compounds.....	199
Figure 7-2: Schematic diagram of a diode array detector (DAD).....	204
Figure 7-3: Examples of calibration graphs obtained	207
Figure 7-4: A representative chromatogram	209
Figure 7-5: Concentrations ($\mu\text{g g}^{-1}$) of eugenol, ME and chavicol	210
Figure 7-6: Concentrations ($\mu\text{g g}^{-1}$) of eugenol, ME, chavicol and methyl chavicol ...	212
Figure 8-1: Different phases of the research study.	215
Figure 8-2: Ser-261 represents <i>EOMT</i> , while Phe-261 represents <i>CVOMT</i>	217

Figure 8-3: AA alignment outputs.	219
Figure 8-4: Polymorphic tables showing several AA similarities and differences.....	220
Figure 8-5: Different cloning strategies..	223

Content of Tables

Table 1-1: Plant families' details containing therapeutic value	18
Table 2-1: Primers used in the study.....	62
Table 3-1: Plant sample details	76
Table 3-2: DNA sample details.....	78
Table 3-3: Biomolecular Technology Group Ocimum Samples.....	79
Table 3-4: PCR programme for <i>trnH-psbA</i> primer pair.....	81
Table 3-5: Biomolecular Technology Group sequences.....	86
Table 3-6: Samples Identification Summary.....	93
Table 4-1: PCR programme for different primer pairs	103
Table 4-2: BTG DNA samples.....	104
Table 4-3: List of accession numbers used throughout the study	105
Table 4-4: The details of various bp-length intron	119
Table 4-5: Polymorphic AA positions	129
Table 4-6: An example of AA polymorphism	130
Table 4-7: AA polymorphism study	133
Table 5-1: <i>EOMT</i> -F-1 and <i>CVOMT</i> -F-1; OMT-R-1074 primers.....	150
Table 5-2: Different thermocycler programmes designed for pET primers	156
Table 5-3: Vector-specific and gene-specific primer list.....	156
Table 5-4: <i>EOMT</i> and <i>CVOMT</i> forward primers	158
Table 6-1: The thermocycler program for <i>Att</i> primers.....	183
Table 7-1: Amounts of estragole in plants essential oils.....	203
Table 7-2: Preparation of different stock solution concentrations	206

Acknowledgements

This research was supported by De Montfort University Scholarship programme. I am really grateful to my supervisors for seeking this funding and supported me throughout an entire journey of my PhD study.

I would like to thank Prof. Adrian Slater for his continuous support from securing the scholarship funding, for his step by step advice during my lab work and for the insightful comments on my manuscripts. I am grateful to Dr. Caroline Howard to come up with this research proposal in Tulsi plant. Her calm and positive approach has inspired me to overcome the failures with no loss of enthusiasm. I greatly appreciate the contribution of Dr. Tiziana Sgamma with whom I was able to share all my ups and downs. Her determination and “can do” attitude help to overcome several hurdles during my journey.

I would like to thank Dr. Eva Masiero for her continuous support and guidance. I am thankful to Dr. Linda Bengtström for her assistance with the HPLC analysis of data. I am sincerely grateful to Dr. Umakanth Venkatraman Girija, Dr. Randolph Arroo and Dr. Ketan Ruparelia for their advice. I am also immensely grateful to other PhD students, project students and frontrunners; it has been a pleasure sharing the lab with them.

This journey wouldn't be possible without the blessings from my parents. I am really grateful to my sister for her never ending support and love. I would like to thank all my wonderful friends to be my stress relievers. I appreciate the efforts of my colleagues at University of Leicester. Their support and encouragement were highly valuable.

Last but not the least, a big thank you to two very important persons in my life, Devang and Kian. Their presence, love, care and encouragement are endless. I am so grateful for their unconditional love and support throughout this journey.

Conference attendance and Presentations

1. The UCL Spices and Medicine Half-day Symposium on Wednesday 11th October 2017, held at the UCL School of Pharmacy followed by Frank Fish Lecture 'The Human Side of Glucose Metabolism' presented by Professor Frances Brodsky, Director of the Division of Biosciences.
2. Allied Health Sciences Research conference 2018 (07/2018). Organised by the HLS department of De Montfort University (Oral Presentation).
3. Trends in Natural Product Research - PSE Young Scientists' Meeting Lille 2017 (06/2017). Organised by the Phytochemical Society of Europe and the Charles Viollette Research Institute. The Young Scientists Meeting entitled Trends in Natural Products Research - PSE Young Scientist' Meeting Lille 2017 Natural Products in Health, Agro-Food and Cosmetics (Poster presentation).
4. Next Generation Sequencing conference (03/2017). Organised by Nucleus Genomics and the University of Leicester.
5. Research and human tissue legislation (09/2016). Medical Research Council. This e-learning module provides an overview of human tissue legislation in the UK; best practice and practical tips for compliance.
6. The 9th Joint Natural Products Conference 2016 in Copenhagen (07/2016). A joint conference organized by University of Copenhagen. The joint conference covered a broad range of scientific disciplines within natural products research. Took part in poster presentation and young researcher workshop during this conference.
7. Annual Health and Life Science Postgraduate Research Conference (06/2016) at De Montfort University. The conference provided an opportunity to showcase and learn about the wide variety of research that is going on in the faculty.
8. School of Allied Health Sciences Research conference (SAHRC 06/2016). Organised by De Montfort University. The event showcased the breadth and depth of the research currently being carried out within the School. There was an

opportunity for oral and poster presentation by postgraduate students, early stage researchers and senior researchers.

9. Professor Frank Fish Memorial Award Lecture 2015 at UCL School of Pharmacy (10/2015). Lecture entitled 'Potential and challenge of traditional Chinese medicine to meet unmet clinical needs' was given by Prof. Tommy (Yung-Chi) Cheng from Yale University.

Abbreviations

£	Pound
\$	Dollar
%	Percentage
µM	Micromolar
µl	Microliter
2,4-D	2,4-dichlorophenoxyacetic acid
AA	Amino acid
Amp	Ampicillin
BAP	6-benzylaminopurine
BLAST	The Basic Local Alignment Search Tool
BLASTn	BLAST Nucleotide
BLASTp	BLAST Protein
bp	Base pair
BTG	Biomolecular technology group
CAM	Commission for Alternative and Complementary medicines
cDNA	complementary DNA
CVOMT	Chavicol O-methyltransferase
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
<i>E. coli</i>	Escherichia coli
e.g.	exempli gratia (“For example”)
EDTA	Ethylenediamine tetraacetic acid
EOMT	Eugenol O-methyltransferase
etc.	Et cetera (“and the rest”)
FRLHT	Foundations for Revitalisation of Local Health Traditions
g	Gram
GOI	Gene of interest
HCL	Hydrochloric acid
Hz	Hertz
i.e.	id est (“that is”)
IAA	Indole-3-acetic acid
IDT	Integrated DNA Technologies
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITS	Internal transcribed spacer
kg	Kilogram
KN	Kinetin
MC	Methyl chavicol
MCS	Multi Cloning Site
ME	Methyl eugenol
MgCl ₂	Magnesium chloride
MHRA	The Medicines and Healthcare Products Regulatory Agency
Min	Minute
MS	Murashige and Skoog's medium
NAA	α-naphthaleneacetic acid
NaOH	Sodium hydroxide

NCBI	National Center for Biotechnology Information
Ng	Nanogram
NIBSC	The National Institute for Biological Standards and Control
°C	Celsius
OMT	<i>O-methyltransferase</i>
PCR	Polymerase chain reaction
PGR	Plant growth regulator
qPCR	Quantitative PCR
RNA	Ribonucleic acid
Rpm	Revolutions per minute
RT-PCR	Real-time PCR
SOC	Super Optimal broth with Catabolite repression
TBE buffer	Tris-acetate-EDTA buffer
T _m	The melting temperature
UK	United Kingdom
USA	United States of America
V	Volt
W/V	Weight per volume
WHO	World Health Organisation
X-Gal	5-Bromo-4-Chloro-3-Indolyl -D-Galactopyranoside

Table of Amino Acids and Their Abbreviations

Amino acid	3-letter abbreviation	1-letter abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1 Introduction

The World Health Organisation (WHO, 2000) has clearly defined the term Traditional medicine that refers to health practices, knowledge, and beliefs based on the theories and experiences indigenous to different cultures. This is a common practice followed by all communities around the world. There are different concepts of health and disease underlying different systems of medicine. The holistic approach of many traditional systems includes the view that a wide range of plants contribute to a healthy life. This explains the large overlap between herbal medicines and the rich use of herbs and spices in traditional cuisines, for example. (WHO, 2005). The concept of modern phytotherapy treatment is evolved from ancient natural therapy. The core knowledge of herb was mainly used according to individual preference (Engebretson, 2002; Conboy *et al.*, 2007; Rishton, 2008; Schmidt *et al.*, 2008).

In the rural areas of many countries, people have retained a knowledge of different traditional herbal treatments to cure certain illnesses. This knowledge is passed on across generations for treatment of variety of mental and physical illnesses (WHO, 2002). Most of the people living in developing countries rely mainly on herbal medicinal products as a primary source of healthcare and traditional medical practice. In the recent times, developed countries have also started taking keen interest in such treatments. Plant based medicines have acquired wider public attention these days. A study by the WHO reveals that almost half of the population in Europe has used traditional medicines at least once in their lifetime, while according to the USA Commission for Alternative and Complementary medicines (CAM), in year 2000, more than \$15 billion was spent on traditional remedies (WHO, 2000).

The majority of cultures were using plant for healing purposes but after the introduction of modern medicine in 19th century, plant-based medicines was considered to be a supportive mode of treatment as countless active compounds have been separated from natural products. Modern medicinal system includes pharmaceutical drugs with scientific evidences gathered from the initial stages. In some

European countries, including Austria and Germany, has developed a wider acceptance towards herbal medicine compared to UK. Unconventional and complementary treatments are being used by many practitioners in Europe. Herbal medicines, acupuncture, homeopathic medicines are the major forms of treatments used by therapist in European countries (Fisher and Ward, 1994).

The majority of developing countries' population have widely been using plant-based medicines for a long period of time. Due to wide acceptance of Plant Science study, this practice is also being carried out in developed countries including US and UK over the last two decades. The reasons for which people choose herbal medicines over synthetic drugs are: they are cheaper alternative, have fewer side effects compared to chemical-based drugs and are part of general practice in communities. The maximum applications of herbal medicine are for health improvement rather than curing life threatening disease (Wachtel and Benzie, 2011). Though, in certain cases, adverse effects have been observed when plant based medicine and synthetic drugs are consumed together (Benzie, 2011).

1.1 Medicinal Plants: Global perspective

Various herbs, shrubs and plants are the essential elements of traditional medicine system and they have been used for treatment of several illnesses (Gurib-Fakim, 2006). Indigenous medical systems in India, China and Africa have traditional plants as main remedies (Kong *et al.*, 2003). These traditional health care systems are different from science-based synthetic drugs treatment. The legacy of such modes of treatments have been passed on from one generation to another by verbal communication (Cotton, 1996). Since centuries, this unique approach has been popular in cultures of Chinese, Indian and African communities. The traditional medicine plants are easily available in nature. They serve as an invaluable source of medicine for people living in rural areas (Roberson, 2008). Due to these reasons, research targeting traditional medicinal plants has getting more interest these days. Jiaxiang's study (Jiaxiang, 1997) revealed that around 5000 plant species (as shown in Table 1-1) from different families contain therapeutic value.

Table 1-1: Plant families' details containing therapeutic value

Sr. No.	Name of the Plant Groups	Number of Species
1	Thalophytes	230
2	Bryophytes	39
3	Pteridophytes	382
4	Gymnospermae	55
5	Angiospermae	
5a	Monocotyledones	676
5b	Dicotyledones	3495
Total		4877

Traditional medicinal plants are processed and used either in liquid forms - essential oils, herbal tea and syrup or in solid form such as rub, ointment, tablet and capsule. These products may contain dry powder or plant extract of the medicinal plant. These plant extracts are prepared by using variety of techniques including hot-water extract, plant maceration involves soaking plant materials in a container, acetic acid extract, alcohol extract and long term boiled extract.

The economics and documentation of traditional plants in each continent is different. In India and China, the use of herbal medicine and associated medicinal plants usage has been documented. Due to this, the majority of traditional plants that have international recognition, come from these Asian countries. Europe and America have started using herbal medicine to create a balance in their diet and lifestyle. In African countries mixed approach has been observed towards traditional herbal medicines (CTA, 2018).

In African continent, the ratio of medical doctors to common people is 1:40,000, while there are 1:500 traditional healers available (Abdool *et al.*, 1994). As people cannot access mainstream doctors, they have no access to prescribed medicines and therefore up to 80% people in Africa, 71% population in Chile and 40% in Colombia rely on the traditional medicines (Shaikh and Hatcher, 2005). Hence, studies show that traditional plant-based medicines are being a very important mode of medication in developing countries.

1.2 Medicinal Plants: Indian perspective

Even though healthcare facilities are readily available in India, plant based medicinal products are widely used by medical professionals who practice traditional medicine concept and follows modern drugs therapy, too. Some Ayurvedic professionals combine traditional and modern knowledge to prepare combinations of plant-based drugs which are then published in journals or books to make them available for the rest of the world (Sai, 2010). Such form of treatment is highly popular in India where healthcare facilities are expensive, and people are inclined towards plant-based medicine which is entirely home based, easily available and comparatively cheaper than other form of treatments (Rajasekharan, 1996).

In the Indian subcontinent, different forms of indigenous systems including Ayurveda, Unani, Siddha, Folk and Sowa-Rigpa, are practiced apart from using allopathy medicine (FRLHT, 2015). Even though healthcare facilities are easily available in majority of the cities in India, plant based medicinal products are widely used by medical professionals.

1.2.1 Ayurveda

Ayurveda is one of the oldest forms of Indian medical practice. References of Ayurveda have been found in literature since second century BC and it is believed to have flourished when schools were running in the form of '*Gurukul*'-an ancient form of education system in Indian subcontinent. Sages started building such small '*Gurukul*' for students to reside in, live a holistic life and study the different elements of nature. Each of these education systems and their founders had a different perspective on Ayurvedic medicine. A group of practitioners believed in concept of personalised medicine believed in studying the habits, genetic factors, body type and then treating the patient, while others believed in starting the treatment and simultaneously observing the changes in the health condition (Rao, 1987). Ayurveda is closely linked with the Hindu religion and its references are found in ancient books of Hinduism. The concept of balanced diet, healthy lifestyle, treatment of various physical and mental illnesses is found in ancient Hindu literatures. All this information suggests close relation of Ayurveda and Hinduism. The Indian ancient literature known as *Vedas* (*Rig Veda*, *Yajur Veda*, *Sam Veda*, and *Atharva Veda*) mention that Hindu God named Brahma,

who is believed to be the creator of the universe, passed on this medicinal knowledge to the founder of *Gurukul*, from where it became available to the society. The script of these literatures was written in “*Atreya*”, an ancient language. In the later years, it was translated and edited and was compiled in the form of “*Charaka Samhita*” which is used by Ayurvedic practitioners till date (Udupa, 1975; Behere *et al.*, 2013). This practice has been existing in Indian society for thousands of years and it mainly constitutes plant-based medicine and specially designed diets. This method of treatment primarily targets the three principles of Ayurveda – *Vata*, *Pitta*, *Kapha* - and imbalance of these three is said to lead to sickness (Joshi *et al.*, 2011). *Vata* energy governs the body movements and breathing activities. Imbalance in *Vata* can lead to thin body structure and sensitive skin type. *Pitta* dosha controls metabolism and digestion. The imbalance of *Pitta* may lead to symptoms like indigestion and mood swings. *Kapha* is associated with human immune system and it also controls cells types in body and therefore its mainly related to immune deficiency sickness (Monier-Williams, 2007). Ayurvedic practitioners try to keep the balance of three principles in human body by prescribing drugs and changing diet types. They strongly believe in the combination of meditation and yoga along with the balanced diet that leads to healthy life span. Each person has unique body type which ultimately leads to build their personality, mood, appetite and body shape. Ayurvedic medication is prescribed after identifying each person’s body type unlike the concept of synthetic medicine that has similar medication for the mass of the population (Telles *et al.*, 2015).

There are eight sub divisions of Ayurveda. ‘*Rasayana tantra*’ is one of the sub divisions which include topics such as rejuvenation, healthy diet program, promoting drugs and health behavior. *Rasayana* are well designed drug recipes that have several benefits including youthfulness, long lifespan, freedom from illness, sharp memory power and strong physique. The recipes of *Rasayana* drugs contain different compositions of plant material among which immunomodulation and anti-oxidant drugs recipes have been extensively studied.

Ayurveda considers the excretion of waste material in the form of urine, sweating and faeces. When three doshas – *Vata*, *Pitta* and *Kapha* are not balanced in human body, the

waste excretion system does not work efficiently which results into several diseases such as indigestion, arthritis, constipation and some forms of skin diseases. Ayurvedic literature also includes the concept of heat (*Agni*) formation in human body which mainly balance digestive system. Imbalance in this system leads to ulcer and other illness (Kurup, 2003). Ayurvedic practitioners consider all these factors including doshas, *Agni* form, excretion system, environment and routine in individual lifestyle before considering any treatment for the illness (Hankey, 2001). The whole ayurvedic system is based on the concept of '*Pancha karma*' for removal of unwanted elements and healing of different body tissues. This has been performed using laxatives, forced vomiting using specially prepared oils, blood purification and administering medicines through the nasal route (Premila, 2006).

Ayurvedic medicines (or Ayurveda) are in practice since several generations and still being used by practitioners who have studied the concept of relationship between nature and elements of human body. In this modern era where synthetic drugs are immensely popular, traditional herbal medicine practitioners need to work hard to shape up the strategies to keep this tradition alive. This is difficult due to very less awareness of basic medicinal knowledge, limited literature and unavailability of enough plant resources. Though, by exploring the roots of this effective plant-based medicine system, a strong foundation for advancement of herbal medicines can be prepared in the modern era.

Ayurveda believes that everyone has their own constitution of body which can be determined by one's surrounding environmental conditions, diet habits and excretion pattern. A combination of plants genomic structure study along with Ayurveda concepts, helps to relate this ancient system with recent trends (Gupta, 2015). This new research stream also emphasises on treating patients from same ethnic origin records. Ayurvedic researchers target uncovered areas of synthetic drug study including personalised medicine, toxic side effects and efficacy issues. Various studies are carried out to know the relationship between ethical background, place where people were born, most frequently visited places and person's BMI index. Each person reacts to the certain treatment according to their body type (Satyamoorthy *et al.*, 2014). Moreover, combination of modern treatments and tools along with ancient Ayurveda concepts can

bring positive results in therapeutic world. Keeping this ancient system alive by applying its principles in individual lifestyle and getting more scientific proofs could convince more people to accept this unconventional form of medicine as a part of their life (Warrier, 2011).

Ayurvedic practitioners can prescribe medicines. Some of these medicines contain both plants and heavy metals (Gunturu *et al.*, 2011). Some studies have revealed serious side effects of these medicines, especially when they interact with synthetic drugs (Chopra and Doiphode, 2002). In some case, core ingredients details and specially, documentation are missing while prescribing ayurvedic medicines, which make them unreliable for human consumption. As this system targets individuals as per their personality type, same medication cannot be prescribed to general mass of people. Each medicine is specially designed according to individual's body tendency. The major drawback is that, plants used in this treatment can be misidentified. Collection from different geographic conditions can have effect on secondary metabolite production. Hence, all these factors may result into a compromise on the plant-based medicine quality standards. (Jaiswal and Williams, 2017).

1.2.2 Siddha

Siddha practice is mainly originated from the Indus valley, the motherland of the Dravidian people. As the habitants of this region migrated to the south region of India around 1500 BC, this practice is very popular within local communities of that region. Along with using plant-based medicines, this practice also believes in some spiritual activities (Thomas *et al.*, 2009).

The primary principle of Siddha is very similar to Ayurveda where they believe in three forms of life existence – Fire, Air and Water. Though, this tradition mainly believes in rejuvenation of life rather than targeted any disease. It focuses on the concept where in human internal energy leads to healthy lifestyle. As it directly targets healing of human soul, it is sometimes categorized as spiritual exercise rather than considering it as medicinal practice. The Central Government's Siddha Research Unit, which is based in Chennai, India, has started running some clinical trials on certain Siddha medicines believed to treat disease including ulcer, hepatitis and many more. (Subramanian and

Madhavan, 1984). Unfortunately, it started on a very small scale and some serious scientific investigations still need to be made.

1.2.3 Unani

Unani, also known as Greek medicine, is mainly associated with *Hakim* tradition which is very popular within the Islam community. It is believed that they practiced the very first form of traditional medicine and this knowledge was spread by Islam communities throughout the world (Abdelhamid *et al.*, 2002). Along with use of basic principles, Unani physicians tried to study different atmospheric conditions, and as a result many new medicinal compositions were documented in the literature. During their migration to India, Unani medicines blended with Ayurveda concepts to design new forms of medicines with original Unani background and adaptation of Indian tradition to firmly established their roots on foreign territory (Ali, 1990). During medieval age, Unani hakims and their medicines names were used as a reference in European countries (Yesilada, 2011). *Hakim Ibn Sina* was most renowned physician who practiced Unani tradition and wrote five volumes of '*The Canon of Medicine*'. This textbook is still used as a main reference by Unani physicians. (Mannan and Kahvic, 2010).

Though Ayurvedic and Unani medicines both target the balance of human body, Unani is different from Ayurvedic tradition which is based on the concept of three doshas, while Unani followers believe that air, water, earth and fire – these four forms help to keep the balance of the human body and imbalance of any of these can cause disease. Unani emphasise four humors (blood, yellow bile, black bile, phlegm) and weather (cold, hot, dry, wet), too (Miraj and Kiani, 2016). Pakistan and other Islamic countries are followers of Unani as they have the highest population of Muslim community who still follow this tradition (Rafatullah and Alqasoumi, 2008). One of the most famous Unani commercial products is *Rooh Afza* (useful to maintain water balance in the body) manufacturing by company called *Hamdard*. This product is widely popular in India and Pakistan and available throughout the world. Mohammed Said, founder of *Hamdard*, had also started academic courses in Delhi, India and Karachi, Pakistan, which is actively involved in Unani medicines manufacturing. (Alter, 2008; Hayat and Naeem, 2014).

1.2.4 Folk and Sowa-Rigpa

Sowa- Rigpa, also known as ‘Art of healing’ or ‘Science of healing’, is one of the oldest traditions flourished in the pre-Buddhist era (Roberti *et al.*, 2011). It is very popular in the Indian subcontinent including Tibet, Nepal and North India (Blaikie, 2013). In addition to the use of plant material and natural minerals, Sowa-Rigpa is based on different behavioral treatments lead by supernatural powers. There is a strong relationship between Buddhism ideals, medicines and astrology. A great combination of Sowa-Rigpa knowledge and modern drugs is beneficial for human health, though it requires proper documentation, in detailed research study and knowledge sharing (Lafitte, 2015). Sowa-Rigpa’s main principles are based on divinations which are not only limited to medicine knowledge but can also predict some of the future events in individual life. One of the most popular treatments within this Tibetan tradition is treating wounds by using pastes of plants and other minerals applied with specially prepared wraps (Chudakova, 2017). Sowa Riga believes in self-healing treatment to connect mind with body, which is said to keep balance in individual behavioral personality.

1.3 Global herbal industry and chequered history of herbal products

Synthetic drugs have been adopted worldwide, yet the majority of developing countries still rely on traditional herbal medicines. Over the past few decades, the herbal drugs industry has expanded in medical as well as economical aspects. In developed countries (US, Australia, Canada and Europe), herbal medicine turnover was around US\$ 30 billion in 2000, which has increased up to 15% in the last decade. US market has started adapting the majority of botanical products and from 1990 to 1997 (Eisenberg *et al.*, 1998) about 380% herbal usage has been recorded. India and China are emerging as leading exporters of herbal products for several countries including Germany, Hong Kong, Japan and USA.

Within the last two decades the pharmaceutical industry has started to take more interest in developing plant-based drugs. Around 25% of prescriptions of US drugs contain plant as a key ingredient. In 1995, 119 drugs and in 1997 about 130 drugs were developed from plants and other natural products. As plant-based medicines demand is

increasing rapidly, it is predicted to rise by 30% over the next decade (Wilkinson, 2000). The majority of these plant-based drugs are imported from China and India as they have enriched traditional medicine history in the form of Traditional Chinese Medicine (TCM) and Ayurveda respectively. Due to increasing demand of plant-based products, these countries are the largest supplier of plants, natural products and herbal medicines. In 2012, China and India exported herbal products worth US\$1329.72 million and US\$ 790.56 million respectively. Apart from these two countries, there are many other countries including Japan, Malaysia and Indonesia also have their traditional medicine share in the international market (Nirmal *et al.*, 2013).

Plants have been used for medicinal purpose from the time civilisation started. The Chinese history records written during 2800 BC show references to herbal remedies. Since then, plant-based products had chequered history during last 5000 years or so. Recently due to some efficacy and side effect issues (including addiction, severe health issues, and even death) of synthetic drugs, herbal drugs market value has been increased as a natural and safe alternative for many diseases (Moss and Yuan, 2006). In the past, many diseases had been successfully treated by using herbal products. Due to open access system, all relevant scientific investigation data are now available. A number of scientific investigations have highlighted the importance and the contribution of many plant families including Liliaceae, Caesalpinaceae, Piperaceae and Asteraceae family plants. These medicinal plants are also useful to isolate active compounds to make new drug composition in pharmaceutical industry.

1.3.1 The Use of Medicinal Plants by Migrant People

In Asia, Africa, Latin America and the middle east, around 90% people are still using plant-based medicines. In European countries, a large amount of population (100 million) use complementary medicines in form of herbal therapy or other treatment. Some European countries, the majority of population use unconventional medicines over modern drugs (WHO, 2011). The herbal medicine market is growing at a good pace. Europe and Asia-Pacific are two fastest growing markets for herbal products where these products are sold as dietary supplements (WHO, 2013).

According to census calculation index in 2013, approximately 200 million humans migrate from their home countries. Over the past 50 years, the number of migrants has grown substantially. There were 79 million migrants reported in 1960, while in 2015 a 200% increase has been reported when this number reached up to 250 million (International Organization for Migration, 2013). Literature indicates how these migrants brought plant medicine knowledge to new countries which affected therapeutic practice within migrated countries social groups (Pieroni *et al.*, 2007). During the time when slavery was practiced, some Africans migrated to Brazil and as a result African traditional plants; *Peperomia*, *Bauhinia*, *Kalanchoe* and *Vernonia* were substituted by similar Brazilian plant genera (Voeks, 1990). One of very interesting aspect arose from study is how the whole migration process affect home-based medicines and how some drug ingredients are substituted when original plant species were not available on the new foreign land. Sometimes, rather than replacing an original plant species, migrants prefer to cultivate it in the foreign land (Suky, 2016). They stay in touch with their friends and relatives living in their homeland from where they could procure plant material along with interesting information regarding their medicinal properties (Volpato *et al.*, 2009). There are various factors that affect how each medicinal plant fit into the new environment. In the absence of formal healthcare, people tend to accept alternative therapy more easily. Social relationship between natives and migrant people have played a vital role in spreading awareness of traditional herbal medicines. Along with all these factors, ease of the rules and regulations in a new country plays a key role in acceptance of herbal plants from other countries (Medeiros *et al.*, 2012).

1.3.1.1 Importance of Tulsi plant among Indian traditional medicine – a good example of people migrating with plants

In Southern part of Asia, Tulsi (*Ocimum tenuiflorum* or *Ocimum sanctum*) plant is very popular within the Hindu community. Tulsi represents a good example of plant that has been moved from one country to another during migrations. The Hindu community highly believes in the healing properties of Tulsi and worships it as God. Tulsi has several health benefits and is used to treat illnesses including skin disease, diabetes and asthma. People prefer to keep this plant in their home and it is one of the widely grown traditional plant within the Hindu community. Tulsi seeds and plantlets were brought to

UK from India and Africa. People also prefer to use their contacts in native countries to imports dry plant material and seeds as Tulsi is difficult to be cultivated in climatic conditions of UK. A combination of religious values and medicinal properties makes Tulsi an outstanding plant species brought by migrant Indians in UK and other countries.

Due to the tropical climate in India, the Tulsi plant is kept outdoors, in the courtyard, and worshiped daily at the time of sunrise and sunset. Generally, different species, dark (Shyam) and green (Ram) Tulsi types are found in Indian household (Kousik and Baldev, 2012). Along with worshipping this plant, they use it as an herbal medicine. There is very little literature available where this information documented (Sandhu and Heinrich, 2005). There is an interesting research carried out by Biomolecular Technology Group (BTG) (<http://dmubtg.wixsite.com/dmubiomoltechgroup>) where one of their research students collected valuable information from Hindu community in UK in the form of questionnaires. The study was designed in the form of series of question based interviews where participants were asked for a sample of their Tulsi plants along with some details including from where, when and how they got Tulsi plant, why they worship it, which plant parts are used as medicine, what knowledge they have acquired from their ancestors, whether they are aware of any interaction of synthetic drugs and so on. (Bhamra, 2016). The interviews helped to understand Hindu community's perspective towards Tulsi, its religious value and health benefits. Social gatherings help to spread knowledge and benefits of Tulsi. This research revealed that Tulsi is one of the valuable traditional medicinal plants within Indian communities which has its own economic, cultural and medicinal importance which encourage people to keep Tulsi plant in their home.

1.3.2 Current status of Indian traditional medicinal plants standards

Since last two decades, more attempts have been made to develop herbal medicine. The changes in drug analysis process and new efficacy standards improve the quality of modern herbal medicines. There are several plant species available which require suitable scientific evidences prior to their introduction to the herbal medicine business (Christenhusz & Byng, 2016). Tropical weather conditions are the source of a large

proportion of the world's recognized medicinal plants and therefore India is the largest supplier of medicinal plants (Seth & Sharma, 2004). Indian traditional medicinal knowledge has been accepted globally, now and standardizations process such as improved quality control can certainly increase their demand in future (Singh *et al.*, 2014). Due to massive changes in quality standards, herbal medicines are allowed in the majority of national pharmacopoeias (Sahoo *et al.*, 2010).

There are many Indian organisations including CSIR (Council of Scientific & Industrial Research), Ministry of AYUSH, IPC (Indian Pharmacopoeia Commission) and IIIM (Indian Institute of Integrative Medicine) that have started working in the direction of improving R&D facilities in India. They are also working towards the quality of herbal raw material so that better quality medicines can be manufactured in future (Rastogi *et al.*, 2015). The quality control guidelines for herbal drug manufacturing have been published by IPC in 2012 (Guidance Manual for Compliance of Indian Pharmacopoeia, 2012) followed by the publication of herbal product monograph and phytopharmaceutical drugs in 2016 (Guidance Manual for Monographs Development of Herbs and Herbal Products including Phytopharmaceutical Drugs, 2016).

1.3.3 Regulation of Indian Herbals

Ayurvedic medicines are widely used throughout the world, but predominantly by the Indian diaspora. (WHO, 2005). According to Ayurvedic textbooks, around 2000 plant species contain useful medicinal properties (Mukherjee and Wahile, 2006). Around 950 plant species are used to make different Indian herbal medicines and their market is widely expanding. The plant- based medicine and research sectors require raw materials, as a result 70% Indian plant material is exported in the form of fresh/dry form to medicine products in other countries. Its net worth is around one million US dollars. Other 30% of exports are covered by natural plant extracts (Government of India, 2007-2012).

1.4 Plant identification methods

Biodiversity issues are rising rapidly due to several unwanted human activities such as growing population, overharvesting and pollution. (Murphy and Romanuk, 2014). Therefore, attempts need to be made to protect the plant species available on the planet Earth (Joly *et al.*, 2014). As a part of this process, plant identification followed by classification into suitable groups (family, sub-family) is essential. The reasons for which plant identification is important are – identifying weed species, examine edible plants quality and checking toxicity level if that yield is using for animal feeding. Plant identification contains a series of processes where species characteristics are evaluated step by step. The identification of a completely unknown species collected from the field requires in depth botanical knowledge. Botanists and specifically plant taxonomists can do identification of plants efficiently compared to amateur plant identifiers (Schmid, *et al.*, 1995). There are different ways to identify plants, morphologically and chemically.

1.4.1 Plant morphology

Morphological identification is based on visible features of different plant parts including leaf shape, aroma, floral structure and fruit composition. The process of morphological identification needs to be followed precisely to narrow down the available options within a targeted family, genus or species. It is always helpful to examine multiple characters rather than rely on single factor to correctly identify the plant species (Molina, 1989).

1.4.1.1 Cotyledon structure

The plants are divided based on cotyledon structure which becomes the first leaf or leaves coming out from the seeds. If there is only one leaf appears from the seed, that plant is identified as monocotyledonous, while the plant having two leaves from the seed are known as dicotyledonous. In addition to this, there are several other characteristics need to be considered to classify plant as monocot or dicot. Monocot plants have parallel leaves veins and floral petals in the multiple of three, while dicots plants have netlike vein structure on the leaves and their petals are arranged in multiples

of four to five. The choice of weed killer is determined by identifying these characteristics. Some weed killing chemicals target monocot plants such as grass; while some target dicot family plants (Ogunmenwo, 2003). For example, 2,4-D is ineffective on most grassy weeds (monocots). This makes it useful in monocot crops, such as grains and turf.

1.4.1.2 Leaf arrangement

There are a number of leaves coming out from the stems nodes, one, two or more than two. Patterns can be identified as alternate, opposite or whorled pattern. Not only that, leaves can be single or compound. There is only one leaf present in single leaf arrangement, while compound leaves contain two subgroups – one known as pinnate where leaflets are arranged in opposite formation, while in a palmate structure, the leaf shape resembles a palm shape leaflet. Above all, leaves can also be divided into subgroups according to their shape. Shapes can be classed as oval, round, heart shape, pointed or sharp. Plants can also be identified based on leaf margin structure which can be smooth or toothy. Apart from this, how leaves attach to the stem can also be considered for identification purpose. Leaves can be attached to the stem with a short or long petiole, while sometimes the petiole is missing, and the leaf is directly attached to the stem. Alternatively, in some plants, the leaves are arranged in a whorl around the stem, which are known as clasping leaves (Ryder, 1954; Waldhoff and Parolin, 2011). Some of the leaf's patterns can be seen in Figure 1-1.

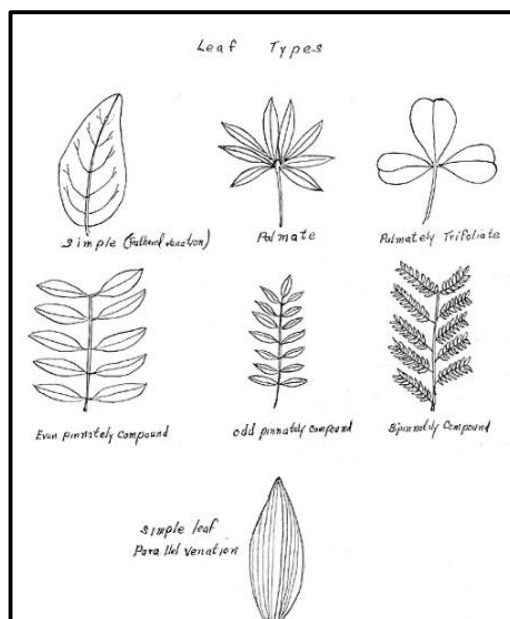


Figure 1-1: Different leaf types: Simple, palmate, palmately trifoliate, even palmately compound, odd palmately compound, bipinnately compound, simple leaf parallel venation
<http://www.kswildflower.org/drawings.php?drawingID=13>

1.4.1.3 Floral structure

The flower arrangements within the plant kingdom are varied and play crucial part in the identification process (Figure 1-2). Moreover, flowers contain four successive whorls which help to narrow down the choices. When a flower is about to bloom and still in the form of a bud, sepal (collectively known as the calyx) protect the flower. Gradually when the bud develops into complete floral structures, petals (collectively known as the corolla) appear. They are attractive flower parts, can be varied in colour and shape. They invite insects for pollination. In some plant species, the sepals and petals cannot be differentiated and are known as ‘tepals’. The male reproductive parts present in the flower are known as stamens, which comprises different structures in the form of anthers (which contains the pollen) and filaments (which support the anthers). The pistil is the female reproductive part of flower. They contain stigma that receive pollen and style which connects the stigma to the ovary (Endress, 2001). Different characteristics such as shape, structure and number of plant parts play an important part in the plant identification process as plants of a same family generally have similar characteristics (Ronse, 2010).

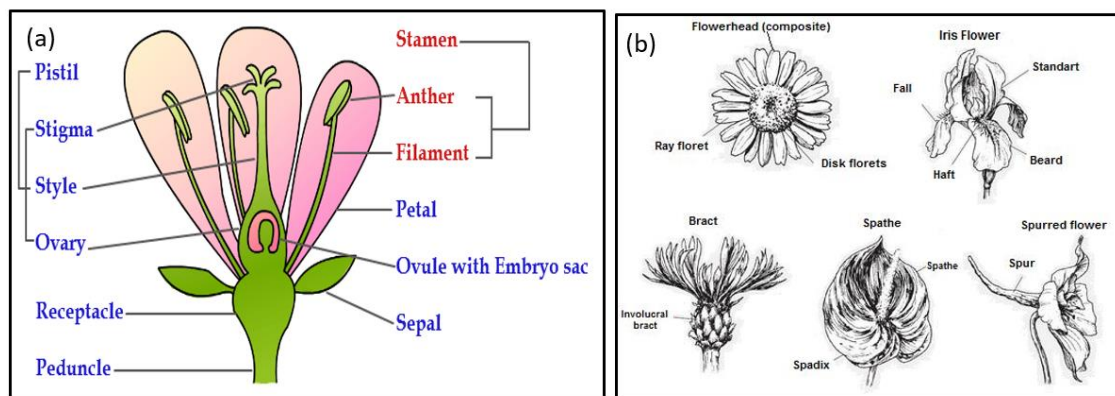


Figure 1-2: Floral structure (a) Basic flower anatomy (Ref.: <https://biology.tutorvista.com/plant-kingdom/flower.html>) and (b) The flower variation (Ref.: <http://www.flowers-gardens.net/gardens/structural-variations.html>)

1.4.1.4 Stem and root system

Stem and root structure can also be included as a tool for morphological identification. Stem shape (straight, square, or cylinder) varies in structure. Sometimes different coloured spots are present on the stems (Gartner, 1995). Similarly, roots can also be divided into different groups. Some of them are described as taproots, which is a single, elongated and dominant root which holds the plant at one spot and is often used for storage of nutrients. Another type is known as the fibrous root system which collects nutrients from the soil and plays an important role in anchoring the plant. In some plants, rhizomes are also present on the roots (Kerk and Sussex, 2012). Ginger, carrot (Khaki *et al.*, 2010) and turmeric (Hoppe, 2011) are well-known examples of tropical and subtropical rhizomes.

1.4.1.5 Seed and fruit morphology

In angiosperms seeds are enclosed within an ovary/fruit; while in gymnosperm, seeds are naked and present on the surface of leaves or scales. In some plants seed and fruit connate (united with each other). Plants can be divided in different categories by examining flower and seed types among which single fruit (develop from single flower), multiple fruit (arise from single flower but multiple pistils) and compound fruit (develop from multiple flowers with single pistil) are more common (Cappers and Bekker, 2013).

1.4.1.6 The Herbarium

The herbarium is a study of collecting and preserving plant species for future reference where whole plant has been dried and pressed by using varieties of tools (Funk, 2003). There are several layers of cardboard sheets used to store these specimens in cupboard for long period of time in alphabetical order (Smith, 2017). The sheet order is decided based on similarities and differences within organisms where closely related and distantly related species are arranged accordingly. Apart from stems, roots, leaves and similar structures (which can be pressed under sheets), other parts including fruits, seeds are stored in preserving jars. This helps to maintain their original structure. Along with storing specimens, herbarium sheets also contain details of collection such as place, time and date. The herbarium sheet is an important tool in plant classification system which allows scientist to work on large number of species. While preparing herbarium sheets, plants are usually collected at the flourished stage (when flowers and fruits develop on plants). Herbarium sheets also contain details such as locality and place of origin of long-lost species which can be used for cultivation purpose.

1.4.2 The classification system

Taxonomy helps to develop classification systems by understanding the evolutionary relationships between different species available on the planet. Swedish-born Carolus Linnaeus is considered to have started using this classification system after publishing his multi-volume book *Systema naturae*, where he described methods of identifying different species and introduce a naming scheme for the classification system (Hodacs, 2010). Earlier, biologists used a very long and complicated naming system. Linnaeus introduced “Binomial” names for each species which contain two Latin names divided into two parts. He travelled all over the world to collect new plant species during his lifetime which helped him to expand his interest in Botany and especially in his classification system. Apart from naming plant species collected by him, he managed to name the species sent by other botanists around the world. Some of them came from newly explored regions of the world (Withgott, 2000; Reid, 2009). Linnaeus had mainly considered using plant reproductive structures for categorisation purposes, which is still considered as the main tool (Sentausa and Fournier, 2013). Some of the names given by him are still been used for plant species that shows the continuing relevance of his

system, which has become the major foundation to understand the complexity of nature (Simpson *et al.*, 2010).

During nomenclature, plants are arranged according to ranking system (Bremer, 2007). Even though this system was designed several years ago, past and modern classification systems are similar in many ways. The majority of taxonomists believed in tree system where genus present in one kingdom is millions of years apart from the genus in another kingdom due to differences evolved during development stages (Sangster, 2014). In the earlier years, living organisms divided into two kingdoms – The living organisms which were mobile and eat other organisms were placed in one group, while plants which cannot move and become food for other organisms were placed in different group. As scientific discoveries happened, classification systems took a better shape and more detailed categories have been added by studying the taxonomical characters of all organisms (Taylor, 1983, Nicolson, 2002). Now, the entire tree of life including plant, animals and other organisms are divided into different kingdoms which are still evolving as a part of natural evaluation process (Cavalier-Smith, 1981).

1.4.3 Chemical analysis

As herbal products usage has widely increased during last decade, pharmaceutical sectors started implementing relevant chemical analysis techniques to check plant compound's quality and efficacy standard. Various analytical techniques are used in drug composition sectors to check quality and adulteration of targeted plants. Different chemical analysis methods are mentioned in the majority of pharmacopoeias (Dejonghe and Russinova, 2014; Gavali *et al.*, 2016).

1.4.3.1 Chromatography

Chromatography technique involves separation using two phases – mobile phase and stationary phase. There are several types of chromatography techniques available as per requirements. All the techniques work on same principle of separating a mixture of chemical substances into its individual components. The testing mixture is dissolved into the mobile phase which travels through the stationary phase to separate individual molecule from the sample mixture. The stationary phase (for example silica beads,

aluminium oxides, cellulose and various reverse phase column) allows molecules to pass through at a particular time. After injecting reaction mixture into the column, separate layers are visible in the form of different colours in TLC (Thin layer chromatography). Each molecule has different adhesion capacity with silica gel and as a result, passes through the stationary phase at different speed. The molecules that bind strongly with stationary phase tend to travel slower, while those that bind loosely, travel faster on the column. Scientists use different types of chromatographic techniques according to their requirements (Pesce *et al.*, 1974).

1.4.3.1.1 *Thin layer chromatography*

In thin layer chromatography, glass plate is coated with a thin layer of silica gel used as stationary phase. The molecules which are more polar, stick to the plate for long period of time while non-polar molecules travel through the plate faster than the polar one. Each component is measured based on their retention time (R_t) which can be found in the literature. While identifying unknown chemicals, the targeted chemical R_t value is compared with the known chemicals. This value is calculated by dividing the travelling time of the target molecule by the travelling time of the total solvent (Hahn, 2006). TLC method is used to separate non-volatile mixtures.

1.4.3.1.2 *Paper chromatography*

Paper chromatographic technique, containing cellulose as stationary phase, is useful for separating coloured molecules, for example pigments. The separated molecules can easily be seen on white paper. Ideally concentrated sample mixture is put on the top of the paper with the help of capillary which then travel gradually on the paper dipped into the solvent. Water, ethanol are used as a solvent those act as a mobile phase while paper acts as a stationary phase (Pazdera, 1966).

1.4.3.1.3 *High Performance Liquid Chromatography*

HPLC technique has widely been used to identify, separate and quantify different component in the sample mixture. A column with chromatographic packing material is used as its stationary phase, while mobile phase is made up of liquid solvent. The sample mixture has been injected which pass through the chromatographic column with solvent mixture under pressure. When sample was injected, it gradually passes through

the column based on its retention time (R_f). There is some trial run carried out with different choices of solvent to get best result. (Töppner *et al.*, 2014).

1.4.3.1.4 High Performance Thin Layer Chromatography (HPTLC)

High-performance thin-layer chromatography (HPTLC) is an enhanced form of thin-layer chromatography. A number of enhancements can be made to the basic method of thin-layer chromatography to automate the different steps, to increase the resolution achieved and to allow more accurate quantitative measurements. High Performance Thin Layer chromatography technique is considered to be a more sensitive technique compared to the other chromatographic techniques which require less sample mixture. In addition to this, it gives sharp and accurate results. This technique has widely been used in pharmaceutical and other industrial sectors to find out the presence of adulterants even when present in very low amount. HPTLC is considered as an expensive technique amongst all other types of chromatography techniques, but the result quality is better and almost accurate by using very less sample material (Srivastava, 2011). It is based on TLC concept, though it allows more accurate semi-quantitative measurements. The main advantages of using HPTLC for detection are scanning, automation, optimization, hypertension and less sample preparation. These features make HPTLC a powerful technique to identify molecules from sample mixture in pharmaceutical as well as research sectors. Some advance features such as modified chromatographic chambers and video scanner makes this technique the best alternative for traditional chromatographic methods including gas and liquid chromatography. This method can produce the results with more accuracy compar to other techniques (Sudberg, *et al.*, 2010).

1.4.4 Plant DNA barcoding

DNA barcoding is a useful tool in various sectors including specie identification where short genetic marker targets the organism's DNA. Along with species identification, it can also be used in ecology to examine animal diet and to find out which plant species consumed by targeted animals (Erickson *et al.*, 2017). Plant barcoding is a useful tool in forensic study when samples are degraded and have lost their main characteristics (Zaya and Ashley, 2012). This system is quite similar to black and white bar-coding marker

system used in the retail operations. This method is not only used for identification purpose, but it is also used to clear taxonomical differences. Though, it is challenging to establish single DNA section in wide plant kingdom the way cytochrome oxidase 1 (CO1) mitochondrial gene has been established in animals as core barcoding region. The search for a plant barcode has involved looking outwith the mitochondrial genome and from the outset many researchers have accepted that multiple markers will be required to obtain adequate species discrimination. Plant barcoding is the method of identifying unknown plant material at species level. Short DNA regions which are commonly present in the majority of species can be determined for selection process (Hebert *et al.*, 2003).

DNA barcoding process is started with collection of specimens which can be obtained from seed bank, open field or green house. Once suitable material is ready for DNA extraction, a tiny sample/tissue is used to carry out series of processes to extract DNA out of it. Once DNA has been successfully isolated, it will be amplified using PCR technique followed by sequencing. The DNA sequence is represented by four letters A (Adenine), C (Cytosine), T (Thymine) and G (Guanine). The resulting sequence is compared with available databases [For example, EMBL (European Bioinformatics Institute), GenBank (National Center for Biotechnology Information), NCBI (National Center for Biotechnology Information)] containing large amount of DNA sequence details. By comparing an unidentified sample with known sequences, species identification process has been successfully completed. This whole process has been illustrated in Figure 1-3 (Meier *et al.*, 2008).

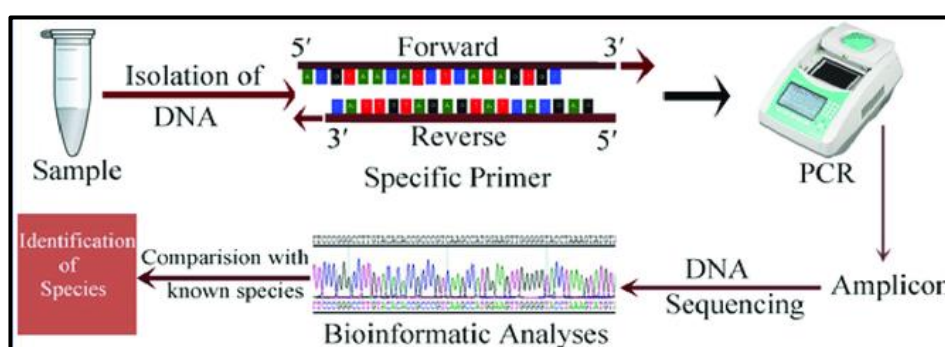


Figure 1-3: DNA barcoding system for identification purpose (Chaudhary and Dahal, 2017)

There are various factors affecting the determination of plant barcode. In the past, numerous projects were designed to evaluate the different techniques used by scientist to design an advance identification of plant DNA barcoding.

1.4.5 Identifying single loci

A single gene fragment (containing 400 and 800 base pairs), which can easily be identified in more than 200,000 plant species can become plant barcode region. It is difficult to target single loci region in whole plant kingdom as plant DNA is much more diverse than animal DNA and scientists accept that multiple marker genes are required for species discrimination. Plant DNA barcoding uses specific regions of DNA in order to identify different plant species. This can be divided into different combinations of markers including *matK*, *trnH-psbA*, and *ITS*. (Hollingsworth *et al.*, 2009; Sgamma *et al.*, 2017). The majority of groups (including the CBOL working group) agreed in 2009 to have *rbcL* and *matK* as the two plant barcodes including *trnH-psbA* and *ITS* as a possible alternative (Seberg and Petersen, 2009; Howard *et al.*, 2011).

1.4.5.1 *matK* and *rbcL* (Maturase K and Ribulose-bisphosphate carboxylase)

The two chloroplast regions recommended by the CBOL (Consortium for the Barcode of Life) are *matK* and *rbcl* (Hollingsworth, 2011). Li and his team (Li *et al.*, 2011) recommended using *matK* and *rbcL* over *trnH-psBA* regions by proposing for Ferns barcodes. *matK* complements *rbcL* with its high sequence variation by providing two locus barcodes. In Figure 1-4, black and grey bars represent ORF (Open Reading Frame) of *matK* and *rbcL* respectively. Some of the relevant primer's positions are represented by arrows on the ruler.

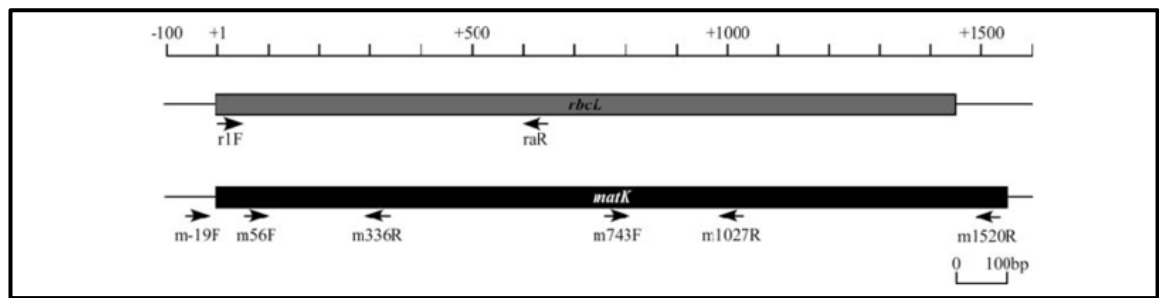


Figure 1-4: Representation of *rbcL* and *matK* gene structure (Ref.: https://www.researchgate.net/publication/223958346_Parentage_determination_of_Vanda_Miss_Jo_aquim_Orchidaceae_through_two_chloroplast_genes_rbcL_and_matK/figures?lo=1)

1.4.5.2 ITS (Internal transcribed spacer)

Apart from plastid regions, the internal transcribed spacers (ITS) region (Figure 1-5) from nuclear DNA would be desirable choice (Howard *et al.*, 2008; Thomas, 2009; Chen *et al.*, 2010). In some plant species, where plastid region has been highly degraded (For example, parasite plants), nrITS (Nuclear Ribosomal Internal Transcribed Spacer) region can be frequently used (Delannoy *et al.*, 2011). Though, there are various reasons that do not allow nrITS to become unique region for plant barcoding purpose. Some of the reasons are: incomplete evolution history leads to different species identification (Möller, 2000), fungal contamination issues (Álvarez and Wendel, 2003), low success rate reported for PCR and poor sequencing result (Gonzalez *et al.*, 2009).

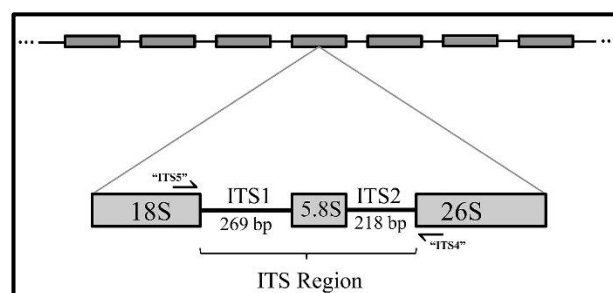


Figure 1-5: Structure of the ITS region (Ref.: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0131185>)

1.4.5.3 Widely recognised regions of plant barcoding *trnH-psbA* (Intergenic Spacer Region)

In the majority of plants species *trnH-psbA* is considered as barcode region (Figure 1-6). It can easily discriminate the different plant species and leads to high success rate

during species differentiation (Shaw *et al.*, 2007). The length of this region can vary from 100-1000 bp. An early termination of some sequencing reads and the repetitive nature of the region makes it difficult to align distantly related sequences. Though by designing new PCR reactions better and clear result outputs can be achieved (Fazekas *et al.*, 2010).

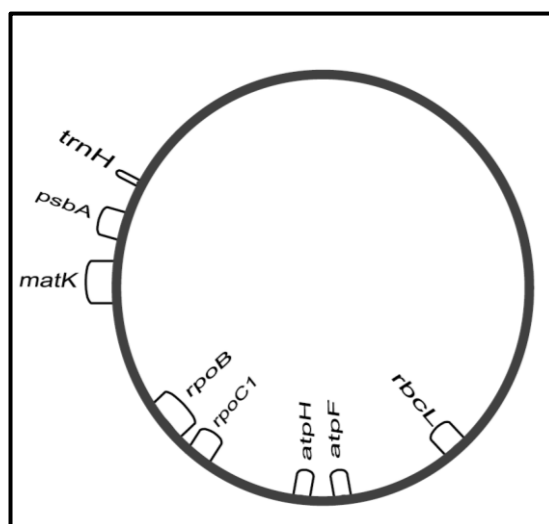


Figure 1-6: Chloroplast genome map showing candidate barcode regions (Howard, 2010)

In this way, hunt for universal plant barcoding region is going on and scientists are unable to establish unique plant DNA region, yet. Though, plant barcoding proved to be an important tool for species identification, in forensic analysis and many other sectors.

1.5 Modern trends in medicinal plants field

The striking difference between traditional medicine and synthetic drug is that former contains plant mixture with several compounds together while later has isolated compounds. Herbal medicines contain a combination of pharmacologically active plant constituents that are claimed to work synergistically to produce an effect greater than the sum of the effects of the single constituents (Ernst *et al.*, 2008). It has been proved that plant extraction mixture shows impressive activity on in vitro culture compare to isolated compounds in equal amount. Synthetic compounds, which are produced in industry or isolated from plants, are less effective as they hardly show same degree of activity compare to crude plant extract (Wagner and Ulrich, 2009). Scientist believes that an absence of interacting elements in modern drugs is responsible for this

phenomenon. Another disadvantage of using synthetic drugs is that they are expensive to produce and therefore unaffordable for poor people, comparatively medicinal plants can be grown anywhere and easily available. (Hirt and M'Pia, 2008).

Along with acting as an active compound, these naturally occurring elements have also earned some popularity as enzymes (to catalyse the chemical reaction) involved in biological processes (Dörnenburg and Knorr, 1995). This attracts scientist's attention to study these plant compounds related genes (Lee and Im, 2012; Gajendrarao *et al.*, 2010). Though, this has given rise to fear of bio piracy (a practice in which indigenous knowledge of nature, originating with indigenous peoples, is used by others for profit, without authorization or compensation to the indigenous people themselves). Bio-piracy is latest trend followed by some big companies to develop improved drugs. They started using develop countries traditional plant medicine knowledge to fulfil their motives and to earn huge profits. As a result, some Indian plants including *Neem*, *turmeric*, *tea*, *tamarind* are patented by foreign countries. Ideally money coming for these inventions is used to improve individual country's infrastructure. Therefore, scientists have started sharing their research study through public domain which could prevent companies to patent it later.

Bio-piracy issue gave rise to Convention on Biodiversity treaty and the Nagoya protocol in 2010 and was enforced in 2014. They work on access and benefit sharing issues of genetic resources available in developing countries. While using these genetic resources, benefits should be shared with the developing countries from where the species originally come from.

1.5.1 Combinatorial biosynthetic technology

Earlier, synthetic drug production and manufacturing traditional herb remedies were two separate branches of medicine which divided communities in two groups – one who firmly believed in modern medicine and other preferred to apply their ancestor's knowledge to treat diseases. In the new era, a mixed approach can be seen where pharmaceutical industries come out to seek new drug compositions from nature rather than relying entirely on industry perspective (Bhatia *et al.*, 1975; Ngo *et al.*, 2013).

Research institutes have started playing an important role as an intermediate step where targeted natural products are first tested according to laboratory protocols. Their efficacy and specificity are examined, and their characteristics can also be modified by using latest biotechnological tools, for example, gene study, pathways details, over expression and gene silencing. (Materi and Wishart, 2007). The difference between developing and developed countries is that the former relies on natural products in their original form while the later use the same compound in form of synthetic drug formation. *Artemisia* plant is an example to prove this fact. The *Artemisia* plant was used by Chinese people to treat malaria long before the Artemisinin drug was developed and started appearing on pharmacy shelves in US, Europe and Asia, (Ram *et al.*, 2014). There are several medicinal plants available from which single or more molecules are extracted from plant. *Rauwolfia serpentina*, which is popular as a snake bite remedy, has been used to produce an antihypertensive agent (Singh *et al.*, 2017); *Cinchona* plant species from the Amazon forest mainly used to treat illnesses including fever is being used to isolate ‘quinine’, a famous antimalarial compound (Uskoković and Grethe, 1973); *Ephedra sinica*, an important ingredient in TCM, is used to produce the ephedra drug for asthma treatment. Though, as mentioned earlier, using plant is different from using single compound, as the plant has many more compounds present which may or may not be active. The remedies provide links for the discovery of new pharmaceuticals and the presence of these compounds in these remedies validates their traditional use.

1.6 Plant and plant material used for the study

As plant is divided into different tissues, only certain plant parts are used for DNA extraction and chemical analysis. Extraction process (both DNA and RNA) has mainly carried out from leaves (young or mature). Depending on plant parts and its usage, DNA extraction techniques and protocols can be designed accordingly (Vijay, 2012). Some old technique including Maceration and manual crushing are still being practiced by small units, while some groups prefer to use modified version of these techniques which is based on microwave, ultrasound and super-particle concept. During DNA extractions, material needs to be handled with care. Each of the plant tissue requires different treatment in the form of suitable chemicals and equipment combination (Li *et al.*, 2013). For example, mangrove plants grow in harsh salt water and to avoid this

adverse condition a thick layer of some special secondary metabolites including polysaccharides, polyphenols chemicals are present within their plant tissues (Bandaranayake, 2002) which hinders DNA extraction procedure. There are several SOPs available in literature to get high yield of DNA without compromising DNA quality and to avoid contamination issues (Vongsak *et al.*, 2013).

For chemical analysis, certain plant parts are collected as some secondary metabolites are only present in particular plant organs. Along with deciding suitable plant parts, temperature and light factors also play an important role while collecting plant material to study secondary metabolites production for chemical analysis.

The types of plant material used in this study were as follows:

1.6.1 Fresh samples

Fresh plant material is fragile and can be dried easily. Therefore, researchers prefer to take some extra precautions while working with the fresh plant samples. As suggested by Sulaiman *et al.*, DNA should be extracted from fresh plant tissues within three hours period after harvesting the plant (Sulaiman *et al.*, 2011). Different chemicals and their quantity used for the extraction process, are also playing important part.

Polysaccharides and metabolites level are relatively low in fresh young leaves (Jobes *et al.*, 1995) and therefore fresh material considered to be more convenient for DNA extraction compare to old material. Due to high moisture content in fresh plant tissue, usage of chemicals during DNA extraction procedure can change compounds' properties.

1.6.2 Dry sample

Though the majority of protocols recommend using fresh plant material, dry material is also using to extract DNA in some cases when samples have to be transported from remote places, or when a plant is only available as herbarium sample. Dry material is more convenient material type to extract DNA. There are some protocols (For example, Qiagen plant DNA) available which can be used for all sample types by altering the few

conditions in beginning including dry samples. Even though, dry plant material is a valuable source of extracting DNA, DNA extracting out of such samples can be highly degraded which hindered the possibility of getting genetic information from an amplifiable DNA. This inability of DNA amplification leads to erroneous sequence information in the later stages.

1.6.3 Commercial products

Apart from using fresh and dry plant material, there are plenty of other commercial products available in market where plant material has been added in different forms. Tablet, capsule, herbal juices, shampoo, tea bags are different commercial products from which DNA can be extracted and analyse DNA sample to check the presence of correct species, contamination and adulterant levels. As each of this product could be highly processed, they require different sample preparation (in the form of clean-up step to remove PCR inhibitors) (Kazi *et al.*, 2013) before starting general DNA extraction process. For example, capsule and tea bags are more similar to dry plant material, but tablet (Chen *et al.*, 2012) and juice (Bai *et al.*, 2013) require protocol modification. In processed material, DNA is highly degraded which makes barcode region shorter than normal and as a result, special PCR test needs to be designed to identify this region during authentication process (Cheng *et al.*, 2014).

1.6.4 Other plant DNA sources

Within the last few decades, significance of different molecular techniques and as a result, preparation and storage of DNA samples has been increased. DNA extracted from the different tissues (chloroplast, mitochondria) can be stored for a long period of time with low cost maintenance facility (Datlof *et al.*, 2017). Storing these samples by preparing in DNA banks in different parts of the world, DNA samples could easily be accessed in future for the majority of purpose such as conservation, training, distribution and different kind of molecular applications (Benson *et al.*, 2000). The majority samples in DNA banks are extracted from fresh plant material, but when live specimens are not available DNA can be extracted from dry material or herbarium samples. (Hodkinson *et al.*, 2007). The concentration and quality of samples in DNA

banks are generally higher, though it can be varied from species to species and person by person who extracted them. The majority of publishers insist submitting sequence details in public databases, but the sequence quality is entirely depending on individual researcher. Therefore, quantification process of DNA sample needs to be optimised before using them for further processes (Harris, 2003).

DNA can be stored in well organised and temperature-controlled environment for several years and in some cases for decades without compromising with quality (Mashima *et al.*, 2017). DNA banks are useful sample storage facilities which require low maintenance, minimum storage facility and can be shared by several researchers by using PCR techniques without repeating the extraction process (Gemeinholzer *et al.*, 2011).

1.7 Secondary metabolites

All plant-based chemicals are results of different metabolic pathways and they have different characteristics in each plant. Plant Metabolites are essential for plant metabolic process. Primary (that is directly involved in normal growth, development, and reproduction of the organism) and secondary metabolites (that is not directly involved in the normal growth, development, or reproduction of the organism) are produced in plants biosynthetically. The entire process involves a variety of enzymes which act on the substrates and as a result these metabolites have been produced. Secondary metabolites are produced from primary metabolites and they do not affect plant growth, development or reproduction stages directly. For example, phenylalanine acts as a primary metabolite which is responsible for the production of eugenol secondary metabolite. Here, phenylalanine is responsible for plant development, while eugenol is mainly being used in fragrance and therapeutic sectors (Ferrer *et al.*, 2008).

Plant's family, genus and species can be classified on the basis of these secondary metabolites group. Secondary metabolites have widely been used in pharmaceutical industry to produce some important drugs for human wellbeing, yet their actual function in plants, is still a mystery. Though some study reveals that due to their toxic nature, they protect plants against pathogens attack. Their production has been greatly been

increased while microorganisms started growing within plant cells. Some secondary metabolites release certain smells which attract predators to kill the pathogens which are harmful for plant development (Taiz and Zeiger, 2006). Scientists are interested to know more about these plant molecules and therefore more secondary metabolites and related research projects are outlining and currently under process.

The majority of *Ocimum* species have been widely studied for its different secondary metabolites present in their essential oils (Wink, 2000). Basil oils from different geographical origins have different features. The composition or quantity can be affected by development stages and other factors during the plant life cycle (Esfahani and Moradi, 2017). Alkaloids, nitrogenous organic compounds, are one of the major classes of secondary metabolites. The essential oils of *Ocimum* contain secondary metabolites in the form of eugenol, chavicol and their respective methyl derivatives: methyl eugenol (ME) and methyl chavicol (estragole) (Jirovetz *et al.*, 2003; Viña and Murillo, 2003). Eugenol and ME have different properties; eugenol protects plants against herbivores (Obeng-Ofori and Reichmuth, 1997), while ME helps in fertilization process by attracting insects (Shukla and Prasad, 1985).

Two chemotypes can be found in *Ocimum tenuiflorum* L. species known as ‘Ram’ (white) and ‘Shyam’ (black) Tulsi. These two chemotypes of Tulsi either have high or low ME:eugenol ratios. As ME is categorised as genotoxic carcinogen, it is important to ensure that the level of these compounds in herbal products fall below the regulatory thresholds (Schinle and Merfort, 2012). As per MHRA guideline, products containing ME have certain regulatory threshold levels limit, for example, maximum 0.01% in fragrance products, 0.0002% in oral cleaning agents, 0.004% in toiletries, 0.001% in skin cleanser and 0.002% in fragrant creams (SCCNFP, 2000; Health Canada, 2010). Urine drug screening of laboratory rodent (National Toxicology Program, 1983) showed different concentrations of ME hepatocytes, while eugenol was absent. Three high doses of ME per week for around 16 weeks triggered liver tumours in laboratory rodents within a year (Borchert *et al.*, 1973; Miller *et al.*, 1983; Wiseman *et al.*, 1987). This high dose of ME was administered by gavage (Long and Jenner, 1963; NTP, 2000). It

would be useful to know more about the genetic determinants of ME levels as that could help to produce plants with a lower ME level (Smith, 2002).

Secondary metabolites pathway steps have been studied in detail until the formation of phenylalanine (Manitto *et al.*, 1974, 1975; Klischies *et al.*, 1975; Senanayake *et al.*, 1977). The large numbers of secondary pathways are made up of the same recurring elements of a few basic reactions. As shown in figure 1-7, the activity of *O*-methyltransferase genes is relevant as they act in the similar way but have different substrates specificities; eugenol *O*-methyltransferase (EOMT) and chavicol *O*-methyltransferase (CVOMT) catalyse the formation of ME and estragole from eugenol and chavicol, respectively (Lewinsohn *et al.*, 2000; Gang *et al.*, 2001).

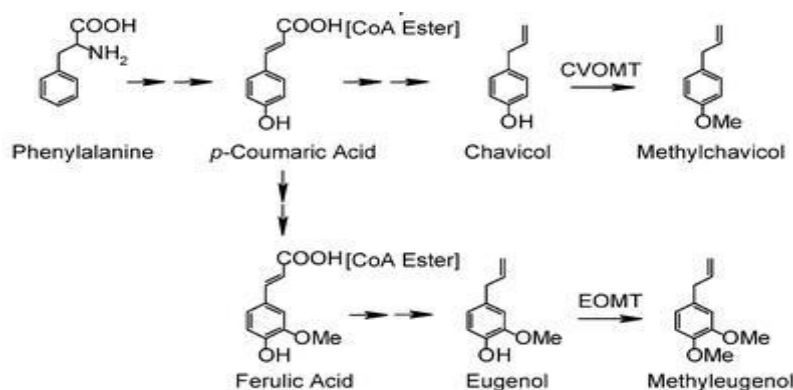


Figure 1-7: Secondary metabolites pathway of Holy basil showing the formation of ME and estragole. Multiple formations are indicated by double arrows, EOMT and CVOMT are enzyme catalysed reactions indicated by single arrows (Gang *et al.*, 2002)

1.7.1 Eugenol (C₁₀H₁₂O₂)

Eugenol is prominently present as an essential oil component of *Syzygium aromaticum* (cloves) and *Ocimum* plants especially in *O. gratissimum* (Anand *et al.*, 2016). Eugenol is a part of the essential oil composition of many different plant species including *Dicopelium caryophyllatum*, *Pimenta dioica* and *Croton zehntneri*. It is a well-known compound in perfume and cosmetic sectors and is commonly used by dentists to alleviate toothache and other dental problems. Eugenol acts as an active biological agent in anaesthesia and to activate the central nervous system as well. (Ueda-Nakamura *et al.*, 2006).

1.7.2 Methyl eugenol (C₁₁H₁₄O₂)

Methyl eugenol is the methyl ether of Eugenol. Along with *Ocimum* plants, it is also present in bay leaves and processed rose oil. Due to strong smell characteristic it has been used as an aromatic agent in detergent, pesticides and perfume industries (Ghosh, 2016).

1.7.3 Chavicol (C₉H₁₀O)

Chavicol (also known as 4-allyl-phenol) characteristics show resemblance to eugenol which is present in basil, bay oil and betel oil. It contains benzene ring with propenyl and hydroxyl groups.

1.7.4 Methyl chavicol (C₁₀H₁₂O)

Methyl chavicol, an aromatic, is mainly present in an essential oil of some plant species including *Ocimum* (Bouvier *et al.*, 2008). This compound has several other names according to IUPAC: estragole, p-allylanisole, 1 allyl-4 methoxybenzene and chavicol methyl ether. Methyl chavicol is used in tomato sauce, perfumes, pharmaceutical products and some beverages. Due to vast usage of this chemical, now lots of research institutes and commercial companies are started taking interest in this plant compound (Martins *et al.*, 2012).

1.8 Lamiaceae family plants

Lamiaceae or Labiatae family (also known as mint family) plants are aromatic and widely used for medicinal purposes. They are either herbs or shrubs, squarish hairy stem, tap -roots and glandular leaves with volatile oils. There are around 240 genera and about 7,500 species. Their leaves are mainly used for culinary purposes. Thyme (*Thymus vulgaris*), mint (*Mentha*), rosemary (*Rosmarinus officinalis*) and lavender (*Lavandula*) belong to this family. The phytochemicals and their phenyl derivatives are well known in the therapeutic world. Methanol, ethanol and chloroform solvents are used to extract secondary metabolites from plants (Pai, 2013). The majority of these plants are cultivated due to their aromatic characteristics and convenient cultivation process. Their leaves are accepted for edible purpose while seeds and stem-cuts are used

mainly for propagation (Raja, 2012). The systematic position of this family plants are as follows:

Division: Spermatophyta

Sub-division: Angiospermae

Class: Dicotyledone

Sub-class: Gamopetalae

Series: Bicarpellate

Order: Lamiales

Family: Lamiaceae

1.8.1.1 Tulsi popularity within South Asian communities

Ayurveda is based on a traditional medical system where Tulsi has special place for its medicinal and spiritual characteristics. This has been proved by scientific research, too. A study revealed that properties of Tulsi plant can address variety of human health problems (Kousik and Baldev, 2012). It can also protect human body against excessive chemical pollutants and help them to maintain their healthy conditions for long period of time. Some study suggests that using Tulsi plant keeps glucose (Suanarunsawat and Songsak, 2005) and cholesterol level (Suanarunsawat *et al.*, 2009) in control and it can also act as an anti-depressant. Its antimicrobial activity suggests it can also be used for sanitising and preservation purpose. Growing Tulsi plants in household atmosphere shows the significance of this plant to protect environment against severe climate change and deprivation issues (Cohen, 2014).

The genus *Ocimum* also belongs to this family and is widely used for culinary and medicinal purposes. This plant is widely spread in different parts of the world among which three different tropical regions of Asia, Africa and America are the main places of origin for the majority of *Ocimum* species (Paton *et al.*, 1999). Crosspollination and certain environmental factors are responsible for differences in plant morphology and secondary metabolites production within *Ocimum* species. The largest number of *Ocimum* species are found in regions of Africa, though certain varieties are also found in the Indian continent as (Mishra *et al.*, 2014). Taxonomists mainly consider

morphological characters while identifying different *Ocimum* species (Conn, 2014). However, very minor differences in morphology at the species level have created considerable confusion. Hence, chemical identification methods of the essential oils of *Ocimum* plants have been considered (Grayer *et al.*, 1996). Though, chemical methods to identify particular species are also problematic as chemical compositions can be affected by various factors including storage, drying, extraction methods, temperature and light sensitivity (Simon *et al.*, 1990). To overcome these difficulties, it has been recommended that chemical and morphological identification along with DNA barcoding methods should be used in parallel (Labra *et al.*, 2004). Chowdhury and his colleagues carried out a comparison study on nine different *Ocimum* species in June, 2017 by using all these three complementary methods and on the basis of this study they correctly identified and classified the nine existing species along with finding a new genotype of *O. africanum* (local name – ‘*Lebu tulsi*’ from West Bengal region, India) (Chowdhury *et al.*, 2017). Its essential oil shows some interesting activity towards gram negative bacteria including *E. coli* (Tamokou *et al.*, 2017).

1.9 The genus *Ocimum*

Ocimum (known as Tulsi or Holy basil), which belongs to the Lamiaceae family and the subfamily Nepetoideae, is an important plant in Hindu community as it is worshiped as a God in daily life and it is commonly known as ‘holy basil’ or ‘Tulsi’. Tulsi is also one of the prominent plants in the pharmaceutical and traditional medicinal sectors. It is a straight, hairy, sweet scented herb. All plant parts including leaves, stems contain useful medicinal properties. Its reference is available in old text books as a ‘magical herb’ which is used regularly as a healing agent (Rastogi *et al.*, 2015), mainly for colds and coughs (Simon *et al.*, 1999; WHO, 2002). It is an important plant, producing high amount of aromatic oil through the trichomes found on the leaf and stem epidermis (Huchelmann *et al.*, 2017). Above all, it is a well-known medicinal plant used extensively to produce a variety of plant-based drugs (Mahajan *et al.*, 2013).

Tulsi’s religious and medicinal status has been accepted in India since several generations. In the recent years, Tulsi’s importance is now being accepted by other countries with scientific evidence. Scientists accept it as a potent drug agent which can

target a variety of diseases in the human body (Cohen, 2014). As stated in Ayurveda, there are three primary forces which are said to balance the human body – Vata, Kapha and Pitta. These forces are responsible for individual physical and mental health. Ayurveda study revealed that Tulsi can successfully balance the Kapha and Vata forms (Singh *et al.*, 2010) and its daily consumption can lead to a long, healthy and stress-free life. Its different plant parts are useful to treat lots of illnesses including fever, malaria, insect or snake bite, headache, cancer, diabetes, asthma, uneasiness, diarrhea and many more (Mahajan *et al.*, 2013; Pattanayak *et al.*, 2010). Tulsi contains high levels of phenolic and antioxidant compounds. Thus, it has been proposed that it can protect the body's cell organelles against unfavorable conditions. For example, during cancer treatment, it can stop DNA impairment and increase cell death by targeting tumor tissues (Siddique *et al.*, 2007; Manikandan *et al.*, 2008). The phenolic compounds can also help the system to successfully excrete unwanted and toxic chemicals by targeting cytochrome P450 enzymes (Rastogi *et al.*, 2007). Researchers are implementing this property of Tulsi plant to against other harmful elements such as pollutants, heavy chemicals and pesticides. When experimental animals are exposed to heavy radiations and then treated with Tulsi plant extracts, result shows reduced numbers of free radicals in cell organelles along with repairing process of damaged tissues (Joseph *et al.*, 2011; Reshma *et al.*, 2012). It can also protect cells against high level of glucose and type 2 diabetic symptoms (Kochhar *et al.*, 2009; Singh *et al.*, 2012).

1.9.1 *Ocimum tenuiflorum* L.

Previously known as *Ocimum sanctum*, is abundant in nature compared to other *Ocimum* varieties. This plant is present in the majority of households of Indian community. It has mainly two varieties: *Ram* (with green leaves) and *Shyam* (with purple leaves) (Raja, 2012). The stem is hairy, erect, branched, scented and around 30-60 cm tall (Figure 1-8). Leaves are toothed, and flowers are purplish or white. Tulsi has been widely spread in whole world as a cultivated plant. It acts as a natural weed killer. According to Ayurveda, *O. tenuiflorum* contains some healing properties. Unani medicine also accepts its magical power and therefore it has legendary status within Indian subcontinent (Warrier, 1995). It is famous for its strong smell and sharp taste. Tulsi plant extracts, Tulsi tea or dry powder are used to cure a range of illnesses including

headache, common cold and cough. Dry Tulsi leaves have also been used to keep insects away from stored grains (Biswas and Biswas, 2005).



Figure 1-8: *Ocimum tenuiflorum*

Tulsi contains variety of phytochemicals and some biologically active compounds. Their characteristics are different within each species and their production is affected by harvesting, storage and processing conditions. All these chemicals are either in seeds (fatty acids, sitosterol) or in leaves (euginal, eugenol, limatrol, carvacrol, estragole) (Kelm *et al.*, 2000, Shishodia *et al.*, 2003). Tulsi's full genome draft has been presented by Upadhyay and his team in 2015 (Upadhyay *et al.*, 2015). They studied transcriptomes of both *O. tenuiflorum* subtypes, *Ram* and *Shyam* Tulsi types.

1.9.2 *Ocimum basilicum*

O. basilicum is also known as 'Sweet basil' (Ronzón, 2012). Its leaves are widely used in culinary world and this is the most accepted edible *Ocimum*. It is almost impossible to identify wild and cultivated basil from its morphology which sometimes creates taxonomical confusion within the genus (Grayer *et al.*, 1996; Lachowicz *et al.*, 1997; Lawrence, 1988). Generally, plant is 35 cm tall with wide, smooth, 8cm long leaves. Flowers are white with purple or red tinged which is around 1 cm long (Figure 1-9).



Figure 1-9: *Ocimum basilicum*

O. basilicum essential oil is trading in market containing some phytochemicals including methyl cinnamate, citral, eugenol and linalool. Their amount can be varied in different basil chemotypes. Simon and his colleagues studied different basil species morphology along with new basil cultivars. A new cultivar known as ‘Lemon basil’ or ‘Sweet dan’ is the combination of all American basil which is fungi (fusarium) resistant, around 70 cm long and contains strong lemon smell. ‘Siam Queen’ is another example of new basil cultivar which has dark green leaves and attractive purple flowers (Simon *et al.*, 1999). An extensive research carried out by Gang and his team on different basil phenylpropanoids including eugenol, ME, chavicol and estragole. They studied these secondary metabolites pathways steps and particularly two enzymes, *EOMT* and *CVOMT*, by using site directed mutagenesis (Gang *et al.*, 2001 and Gang *et al.*, 2002).

1.9.3 *Ocimum gratissimum*

Ocimum gratissimum, which is also known as Vana, Verenda (Raja, 2012) or alfavaca (Silva, 2005), is mainly grown in warm-tropical regions. This is a perennial, woody shrub with simple or branched flowers. Its leaves are wide, dark green, contains aromatic odour and bitter in taste (Figure 1-10). Tannin, terpenoid and alkaloid phytochemical compounds are present in *O. gratissimum* leaf extract (Gupta *et al.*, 2011). The plant parts of this species are used to treat variety of illness including snack bites, fever, cough, pneumonia and some skin disease. *O. gratissimum* plant extract shows some antifungal activities as well (Silva, 2005).



Figure 1-10: *Ocimum gratissimum*

There are some of the major compounds including eugenol, cirsimaritin are presence in *O. gratissimum* essential oil (Vieira *et al.*, 2001). This finding regarding eugenol production within *O. gratissimum* is later supported by Anand and his team (Anand *et al.*, 2016) where they carried out parallel study with different *Ocimum* species among which the majority of eugenol production has been observed in *O. gratissimum* varieties.

1.9.4 Other *Ocimum* species

Apart from these three major *Ocimum* species, there are plenty of other species growing in different environment conditions: *O. citriodorum*, *O. klimandscharicum*, *O. selloi*, *O. americanum* and so on. Each of these species has diverse chemical compositions due to different genotypes. The common usage of these varieties is to extract an essential oil which plays an important part in medical world as a therapeutic agent (Pandey, 2014).

1.10 Phylogenetic inference: concepts and methods

Phylogeny information is illustrated as a tree or branch structure with the help of nucleotide or protein sequences. Earlier only evolutionary, ancestry or taxonomy fields were using this tool, but now it is highly popular in number of other sectors including molecular biology, plant biotechnology and bioinformatics. Along with the wide spread use of sequence analysis, phylogenetic study become main tool to identify relevant species family along with inter-species relationship (Choudhuri, 2014).

During tree formation branches represents roots relations with different species present at the tip of each branch. Based on this roots and branches relationship, species identification and specially its evolutionary history details can be recognised easily (Silvertown *et al.*, 2006). When tree shows more than two root lines, first line shows common ancestors for all species included in study while all lines afterwards illustrate the majority of common ancestor for selected species. In other words, species can be more related to each other if they have a common ancestor and vice versa. It is difficult to establish any relationship between species having different ancestors as branch does not show any time-period during phylogenetic tree formation (Swenson, 2011). The physiological, chemical and morphological characteristics are helpful to build phylogenetic tree. These characteristics can either be ancestral, coming from same ancestor or evolve gradually over many generations (Choudhuri, 2014). Big variances can be seen among less related species while closely related species show fewer differences.

1.11 Research aims and objectives

India is one of the largest suppliers of medicinal plants to the world. Officially around 3,000 medicinal plants are documented lawfully, yet practitioners are believed to use more than 6000 plants species for medicinal purpose (Sai, 2010). There are various databases available that classify medicinal plants in different categories for different remedies. Knowledge and collaboration projects of different traditional system could certainly bring astonishing results and prove beneficial for humankind. Foundations for Revitalisation of Local Health Traditions (FRLHT) hosts a database containing botanical and vernacular names along with authenticate images of each plant species. They try to classify and distribute all the plant species in different traditional systems.

Initially the aim of this project was to use a rational approach to determine the best target plant from the list of interested traditional Indian plant species (Figure 1-11) with outstanding medicinal properties.

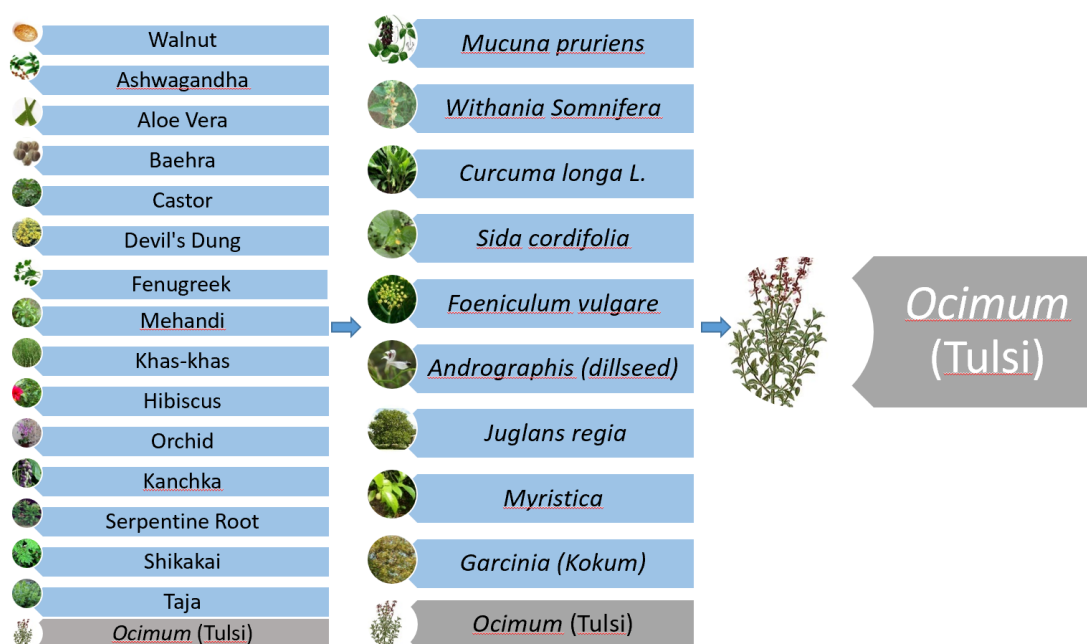


Figure 1-11: An overview of beginning of the research project containing Indian Traditional Medicinal plant list

While studying selected Indian traditional medicinal plants characteristic, Tulsi was one of the prominent plant with the subject of numerous therapeutic and pharmacological applications within several scientific studies. Tulsi has various pharmacological actions including antimicrobial, immunomodulatory, adaptogenic, antidiabetic, anticancer and many more. Tulsi leaves are commonly used plant parts known to have several bioactive compounds including eugenol, linalool, ursolic acid and 1,8-cineole. Eugenol is the major compound present in Tulsi varieties though each of cultivar contains different amount of eugenol produced in their glandular trichomes. Recently, eugenol has been acquired attention as antidiabetic molecule, though Tulsi phytochemical composition is very complex which depend on different factors. Along with eugenol, there are some other phenylpropanoids including methyl eugenol, estragole, rosmarinic acid are also present in Tulsi's essential oil (Jamshidi and Cohen, 2017). These phenylpropanoids have different characteristics that are different from eugenol and other compounds. As Tulsi contains diverse secondary metabolites compositions, out of all other selected Indian traditional medicinal plants, it has been chosen as main plant for this research study.

Based on above findings, previous related research articles have been studied to get more details about secondary metabolites presence in Tulsi's essential oil. Tulsi plants are characterised by high levels of essential oils containing eugenol, methyl eugenol and estragole (methyl chavicol). Two chemotypes of Tulsi have been distinguished, based on high or low methyl eugenol:eugenol ratios (Schinle *et al.*, 2014). As methyl eugenol and methyl chavicol are classed as genotoxic carcinogens, it is important to ensure that the levels of these compounds in herbal products fall below the regulatory thresholds. The levels of methyl eugenol in *O. tenuiflorum* are generally higher than in other *Ocimum* species. The conversion of eugenol to methyl eugenol is catalysed by *eugenol O-methyltransferase (EOMT)*. *EOMT* gene sequences have been isolated from a range of *Ocimum* species and from different chemotypes of *O. tenuiflorum*. Analyses of *EOMT* genomic and cDNA sequences revealed a 843 bp open reading frame and the presence of a 90-104 bp intron (Gang *et al.*, 2001; Gang *et al.*, 2002; Renu *et al.*, 2014).

Based on these literature studies, full length *EOMT* cDNAs were prepared from *O. tenuiflorum*, *O. gratissimum* and *O. basilicum* plants and used for bacterial transformation. The levels of 4 secondary metabolites (eugenol, methyl eugenol, chavicol and methyl chavicol (estragole)) in these three species were also determined by HPLC results. The long-term aim was to determine the effect of *EOMT* transgenes on the levels of these metabolites and to fulfill that purpose four binary vectors were prepared for future plant transformation work. In brief, the aim of this research work is to determine the genetic basis of high and low levels of ME in different chemotypes of *O. tenuiflorum* (holy basil or Tulsi). The key objectives for this work are divided in to two major parts.

1.11.1 Explore the occurrence of DNA sequence polymorphisms within the key biosynthetic genes for ME and estragole pathway

The main aim of this project is to increase understanding of the genetic basis of species variation in the profiles of medicinally significant secondary metabolites such as eugenol, ME, chavicol and estragole.

1.11.2 Determining the effect of genetic differences on enzyme activity and substrate specificity

Bacterial transformation tool was explored to study the production of relevant proteins. This was determined the effect of genetic differences in protein sequences on enzyme activity.

2 General materials and methods

This chapter outlines the common materials and methods used throughout the project. Specific protocol details will be described later in the thesis in relevant chapters.

2.1 Plant Material collection and storage

2.1.1 *Ocimum*

Fresh plants, twigs and seeds of different *Ocimum* species were donated by members of the Hindu community from the UK (from several locations across UK including Leicester, Preston, Milton Keynes and London time to time), Egypt (in July'14) and India (from state Gujarat in Jan'15). Some seeds were also obtained from online seed suppliers; Jungle Seeds (www.jungleseeds.co.uk) and Chiltern Seeds (<https://www.chilternseeds.co.uk/>). Dry powder and plant material collected by a previous De Montfort University PhD student were also used for this study (Bhamra, 2016). Fresh basil plants and seeds were collected from Sithonia, Greece (in Sep'16). Samples of home grown Tulsi plants were collected from members of local communities. Seeds and plant material from a range of *Ocimum* species and chemotypes were provided by Professor Peter Nick, Karlsruhe Institute of Technology, and Dr Eike Reich, CAMAG, Switzerland.

2.1.2 *Arabidopsis*

Arabidopsis seeds: N1092 – Columbia wild type (Col-O), N502373 – Columbia wild type with T-DNA insertion (OMT) and N646526 – OMT mutant type were obtained from the Nottingham *Arabidopsis* Stock Centre (NASC).

2.1.3 Prior procedure on samples before extracting DNA

Samples were obtained either as a fresh or dry material and they were further processed in the lab to homogenise the tissue. Dry samples were stored in control environment until crushed using a TissueLyser II (Qiagen Inc., Germantown, MD, USA) and tungsten beads. The crushed material was used for DNA extraction as described in

section 2.3. Fresh samples were crushed with the help of lysis buffer and 4µl RNase A. Samples of fresh material were stored at –80°C after harvesting.

2.2 Seed germination

Ocimum and *Arabidopsis* seeds were sown using different techniques, but the seeds surface sterilization process performed was the same. The seed surface was sterilised with a combination of 20% SAVLON Antiseptic Cream (GlaxoSmithKline Consumer Healthcare - UK) and few drops of TWEEN 20. Seeds were soaked in this solution for 10-15 mins followed by 3-4 wash using double distilled water (DDW).

2.2.1.1 Filter paper germination

Petri dishes with damp filter paper (Whatman® qualitative filter paper) were used for small size seed germination. All tools used were previously sterilised with 20% Virkon or 10% bleach. Each dish was labelled with the seed's name, number and date on which they sown. The paper was soaked with DDW and seeds were placed on the paper. Seeds were spread evenly. Petri dishes were secured with parafilm which prevents the evaporation of water whilst allowing gas permeability. Dishes were kept at 24-26°C. Germination time was recorded.

2.2.1.2 MS medium germination

A combination (in amount for 1 liter solution) of 4.3g MS medium (Cat. No. - MD154, Phygenera®), 6g (0.6%) agar (Agar No. 4 Plant Tissue Culture Grade Product Code: MC029, Lab M™), 30g (3%) sugar (Cat. No. - S0389, Sigma Sucrose) and DDW was autoclaved and poured in petri dishes. Surface sterilise seeds were placed on this solid medium and dishes were secured with parafilm. The same medium placed in magenta jars was used to grow plantlets after germination.

2.3 Genomic DNA extraction

Genomic DNA was extracted from either 100 mg of fresh material or 20 mg of dry material using the DNeasy Plant Mini Kit (Qiagen Inc., Germantown, MD, USA)

following the manufacturers' guidelines. Dry material was homogenised with a TissueLyser II (Qiagen Inc., Germantown, MD, USA) using a tungsten bead placed in a round-ended microcentrifuge tube (30Hz for 1 min).

Fresh plant samples were first manually disrupted to a fine powder in liquid nitrogen with a mortar and pestle. A lysis solution comprising 4µl 100 mg/ml RNase and 400µl of AP1 buffer (lysis buffer) was added to the sample and all remaining steps were carried out as described in the Qiagen DNeasy Plant Mini protocol (DNeasy® Plant Handbook pp22-25, Qiagen, 2012). At the final step, 100µl of buffer AE was added to the DNeasy spin column membrane. It was incubated at room temperature for 5 minutes followed by centrifugation at 8000 rpm for one minute. This step has been repeated to prepare three DNA aliquots for short-term storage at 4°C (100µl), medium-term storage at -20°C (50µl) and long-term storage at -80°C (50µl). DNA quality and quantity were checked with the NanoDrop™ ND-1000 Spectrophotometer (Thermo Scientific).

2.4 RNA extraction and cDNA synthesis

The RNeasy Plant Mini Kit (Qiagen Inc., Germantown, MD, USA) was used to extract RNA, following the manufacturers' guidelines, from fresh *Ocimum* plant samples grown in the De Montfort University greenhouse.

RNA quality and quantity were checked with the NanoDrop™ ND-1000 Spectrophotometer (Thermo Scientific) and 1-2 µg of total RNA was run on a 1% w/v agarose gel (with 1X TBE buffer at 90 voltage for 30 mins). Genomic DNA was removed with a TURBO DNA-free™ Kit (Thermo Scientific - Catalog number: AM1907) followed by a PCR run with the specific primers: Renu EOMT-F and Renu EOMT-R, (see Table 2-1), to prove the absence of DNA contamination. As this primer is *OMT* gene specific, negative PCR result shows an absence of genomic DNA and clean RNA product. PCR was performed as described in section 2.5 and PCR products were visualised on gel as described in section 0.

Bioline SensiFAST cDNA synthesis kit (Bioline SensiFAST cDNA synthesis kit, Cat. No., BIO-65053) was used to synthesise cDNA from 3 µg of total RNA following the manufacturer's guidelines.

2.5 Polymerase Chain Reaction (PCR)

The standard PCR master mix contained the following components: MyTaq Red mix (2x, Bioline), 0.2 µM of each forward and reverse primer, DNA template (typically 2µl of plant genomic DNA extraction, 1µl of cDNA synthesis reaction or 1µl from a touch of a single bacterial colony dispensed in 50µl DDW), made up to a total volume of either 10, 20 or 50 µl with DDW.

The amplification consisted of an initial denaturation at 94°C for 2 min, followed by 30-40 cycles (as specified) of denaturation at 94°C for 30 s, annealing (at specified temperature) for 30 s, and extension at 72°C for 45 s per kb of expected product. A further 2 minutes of extension at 72°C was carried out at the end of the cycles. Primer sequences, annealing temperatures and expected product sizes are presented in Table 2-1.

2.5.1 Primer design

Table 2-1: Primers used in the study

Primer Name	Sequence (5'-3')	Annealing temp.	Product size
<i>trnH</i>	CGCGCATGGTGGATTCAATCC	Touchdown temperature begin at 58°C, reduce by 1°C per cycle until 48°C	500 bp
<i>psbA</i>	GTTATGCATGAACGTAATGCTC		
Renu <i>EOMT</i> -F	TGTCGACAGAGCAACTTCTT	55°C	843 bp
Renu <i>EOMT</i> -R	GGATAAGCCTCTATGAGAGACC		
Int-F	ACGACATGTGGGAGGTCAAT	53°C	204 bp
Int-R	GCCGAGCATGGGTAGATTTT		
<i>EOMT</i> -F-1	ATGGCATTGCAAAAAGTAGA	52.6°C	1074 bp
<i>CVOMT</i> -F-1	ATGGCATTGCAAAATATGGA		
<i>OMT</i> -R-1074	TTAAGGATAAGCCTCTATGAG		
pET_ <i>EOMT</i> -F	AAAAAAGCTAGCATGGCATTGCAAAAAGTA	Initial 5 cycles of denaturation at 48°C followed by 30 cycles of denaturation at 58°C	1102 bp
pET_ <i>CVOMT</i> -F	AAAAAAGCTAGCATGGCATTGCAAAATATG		
pET_ <i>OMT</i> -R	TTTTTTGAATTCTTAAGGATAAGCCTCTAT		
Att_ <i>EOMT</i> -F	GGGGACAAGTTTGTACAAAAAAGCAGGCTA AAAAATGGCATTGCAAAAAGTAGATATTT	60°C	1146 bp
Att_ <i>CVOMT</i> -F	GGGGACAAGTTTGTACAAAAAAGCAGGCTA AAAAATGGCATTGCAAAATATGGATATTT		
Att_ <i>OMT</i> -R	GGGGACCACTTTGTACAAGAAAGCTGGGTTT AAGGATAAGCCTCTATGAGAGAC		

2.6 Gel Electrophoresis

Gel electrophoresis was used to separate DNA bands based on their sizes. PCR products, cDNA and plasmids were loaded on a 2% (w/v) agarose gels prepared with 1x TBE buffer (made up of 10.8 g Tris base, 5.5 g Boric acid and 0.584 g EDTA for 1 litre) with 2µL SYBRsafe™ DNA stain (Invitrogen, Cat. No. - S33102 SYBR™ Safe DNA Gel Stain).

RNA was checked by running 1µl of each sample on a 1% (w/v) Agarose gel prepared with 1x TAE buffer and 2µL SYBRsafe™.

Loading buffer (Bioline - 5x Loading Buffer Red BIO – 37068) was used when necessary.

A DNA ladder (Bioline Cat. No. - BIO-33045 EasyLadder 1) used to determine the size of DNA fragments. Electrophoresis was carried out at 90V for 25-30 minutes in general (voltage and time limit were changed according to expected product size). Agarose gels were then analysed with a Gel Doc™ EZ Gel Documentation System (BioRad, Oxford, UK).

2.7 Isopropanol Clean Up

Further DNA clean-up was performed when PCR reactions gave negative results using universal primers such as the *trnH-psbA* barcode primers. DNA, usually 50µl, was mixed with 35µl of chilled isopropanol and centrifuge at 11,000 rpm for half an hour at 4°C. The supernatant was discarded, and pellet was re-suspended in 200µl of 70% cold ethanol. The sample was then centrifuged at 11,000 rpm for 10 mins at room temperature. The resulting supernatant was discarded, and the tube was kept open for half an hour to let the ethanol evaporate. The resulting pellet was then dissolved in 50µl TE buffer (from Qiagen DNeasy Plant Mini Kit) and used in PCR reactions as described in section 2.5.

2.8 Purification of PCR products from gels

If multiple bands were detected after electrophoresis, the correct size bands were cut from gels and purified using a QIAquick Gel Extraction Kit (QIAGEN, Cat. No. 28704) following the manufacturer's guidelines. Purified products were eluted in 30 µl of DDW.

2.9 PCR purification

PCR purification used as an alternative to gel purification technique when only one band was detected on the gel. Samples were purified by using Qiagen PCR purification kit (Cat No./ID: 28104, QIAquick PCR Purification Kit). The purified product was collected in elution buffer, the quality of sample was quantified using nanodrop and analysed by running on the gel with the help of loading buffer. The samples were either sent for sequencing or used in other steps.

2.10 Bioinformatics tools

Positive 50ul PCR reactions were sent for DNA sequencing to Macrogen (<http://foreign.macrogen.com/eng/>) in The Netherlands. The result output comes in Sanger sequencing form along with GC% and trace data. Chromatograms were analysed and edited using CLC Bio Main Workbench (Workbench 7) (<https://www.qiagenbioinformatics.com/products/clc-main-workbench/>).

DNA sequences were trimmed, and different reads of the same template were assembled into contigs using the default settings in the CLC platform. Where necessary, contigs were created by relaxing the stringency of the settings and/or using a reference sequence. Consensus DNA sequences were aligned, and the correct reading frame was converted into an amino acid (AA) sequence.

NCBI BLAST tool was used to analyse the experiment results and to find reference sequences for nucleotides (BLASTn – Nucleotide BLAST) and AA sequences

(BLASTp – Protein BLAST) using the default settings. MegaBlast search tool was used for quick DNA search.

An algorithm (for example UPGMA), multiple alignment programs (for example T-Coffee) and visualisation apps (for example Jalview) were used to prepare phylogenetic tree.

2.10.1 Gel Purification

While running restriction enzyme products (both vector and Gene of Interest (GOI), identification of correct size bands is absolute necessary. This can be done by using a suitable DNA ladder marker. Along with running PCR products, uncut genes and plasmids can also be run alongside as a control. Use of wide toothed combs and skipping alternate wells are troubleshooting options to see the product more clearly on the gel. 1 volume of DNA loading buffer (Bioline - 5x Loading Buffer Red BIO – 37068) was added to four volumes of sample for easy recognition.

Around 50µl of each PCR product was run and visible bands were cut from gel. The purification of each band was carried out by using the QIAquick Gel Extraction Kit (QIAGEN, Cat. No. 28704). The final product was collected in 30µl of double distilled water and DNA concentration was checked by Nanodrop instrument.

3 Authentication

3.1 Introduction

There is a huge market for medicinal plants which are used for pharmaceutical products and herbal medicines. Due to an increase in the demand for these plants, it is difficult to achieve accurate and rapid authentication of actual plant species in trade and their adulterants. Hence, a powerful and practical tool is required for identification of correct plant species as well as their adulterants for ensuring safety of herbal medicinal products at the scale of international trade.

Plant identity can be confirmed by a number of tests within the industrial as well as the research environment. Both macroscopical and microscopical features are examined at the first stage of morphological identification (Sahoo and Manchikanti, 2010). During second stage of testing, a range of analytical chemistry methods are employed.

Pharmacopoeia and monograph recommend using TLC (Thin Layer Chromatography) technique as one of the standard analytical methods. This technique's standard can be further improved by the development of HPTLC method. Though, chemical markers used for identification purpose have been found in similar plant species within the same genus (Booker, *et al.*, 2016). DNA barcoding method is intended to be complementary to current chemical and morphological identification tests. DNA barcoding has been applied to the identification of plant specimens among herbal drug products within the industrial quality control sector (Sgamma *et al.*, 2017). An industrial approval of this DNA based identification method has been gradually accepted by regulatory standards and monographs as well. Along with identifying medicinal plants, DNA-based methods are also useful for herbal products authentication. Recent reviews have highlighted the importance of the DNA barcoding methods for the medicinal plant industry (de Boer *et al.*, 2015; Mishra *et al.*, 2016).

3.1.1 DNA barcoding

Morphological identification is an ancient and traditional method for species determination. This kind of identification depends on the subject experts but sometimes even experts encounter specimens that are difficult to identify. DNA barcoding offers an alternative approach to this problem. This is an easy and convenient method for species level identification. DNA barcoding is a combination of taxonomy, genetics and bioinformatics that helps not only to identify unknown species, but it is also used to tell species apart as it acts as a unique identifier for individual species (Hebert, 2003). DNA barcodes are short regions (between 400 and 800bp) that can be easily sequenced for each species (Savolainen *et al.*, 2005). An increasing number of standard sequences, uploaded by scientists from all around the world, are available in online databases such as GenBank and BOLD, to which sequences from unidentified samples can be matched.

The DNA barcoding region must contain similarity and variability at the same time. The similar region is required for the use of universal primers, while variable region is useful to be able to distinguish between species (Rai *et al.*, 2012). The Plant DNA has been mainly stored in the mitochondria, chloroplast and nucleus regions. Cox1 (C oxidase gene) from mitochondria, *matk-rbcl*, *trnH-psbA* from chloroplast and ITS (The internal transcribed spacer) from nucleus are used as DNA barcoding markers in the plants. Though due to its low level of sequence variation, Cox1 region is ineffective for authentication of plants.

CBOL (The Consortium for the Barcode of Life group) is an international group dedicated to supporting the development of DNA barcoding. Initially various combinations of plastid markers were proposed (*rpoC1+rpoB+matK* or *rpoC1+matK+trnH-psbA*; *atpF-H+psbK-I+matK* and *rbcL+trnH-psbA*) as the plant barcode (Chase *et al.*, 2007; Kress and Erickson, 2007). Eventually, the *matK* and *rbcL* regions from the plastid genome were selected as “core barcodes” for plants (Hollingsworth, 2011). Along with these two chloroplast coding regions, one non-coding chloroplast region, *trnH-psbA*, has also been identified as an efficient option for sequencing (Vassou *et al.*, 2015). Short length, interspecific variation and good priming sites, make *trnH-psbA* a strong candidate for barcoding (Kress *et al.*, 2005; Howard,

2010). Having once been discounted, the nuclear Internal Transcribed Spacer (ITS) region was again shown to be an effective barcode (Chen *et al.*, 2010) as it has evolved quickly and is useful to determine interspecies as well as intraspecies relationships (Cullings and Vogler, 1998). Though the CBOL group has only approved the ITS region as a supplementary barcode due to some of its limitations (CBOL Plant Working Group, 2009). It is difficult to amplify ITS regions due to development of secondary structure which results in poor sequence quality. Low PCR success and multiple copy numbers have also raised some concerns. In cases where plants contain fungal endophytes, identification can be challenging with ITS alone as ITS is the most commonly used region in fungal identification (Álvarez and Wendel, 2003). All these factors have reduced the efficiency of ITS region (Kress *et al.*, 2005). For these reasons the ITS region was not selected by CBOL as a core plant barcode.

In this way, there has been a considerable discussion about the choice of DNA barcoding region. The use of one or more regions as plant barcodes is of proven value in the molecular biology world, although the idea of selecting one region which fulfils the requirements for all plants appears unattainable. Henceforth, selecting the most appropriate region on a case by case basis is much more effective rather than selecting single region for entire plant kingdom.

3.1.2 Advantage of using DNA based identification techniques

The production of herbal medicines requires raw material from medicinal plants. As the usage of plant based medicines has become widespread, the importance of developing the plant identification methods by which genuine raw materials can be identified, has increased, too (Kazi *et al.*, 2013). Morphological identification techniques cannot be used on processed plant material and as a result new or additional techniques are urgently required in the industry (Poornima, 2010; Newmaster *et al.*, 2013). Traditional methods of species identification require scarce professional knowledge while chemical profiling techniques are unable to provide effective results if the material is affected due to poor storage conditions, physicochemical reactions or is from a closely related species.

DNA authentication has several advantages over morphological and chemical techniques. Plant morphological features can be changed or degraded due to processing, harvesting and storage techniques (McCutcheon, 2002), while DNA barcoding is not restricted by these factors. Phytochemistry is expected to vary between young and mature plant material, while DNA analysis is not affected by plant age (Kazi *et al.*, 2013). Moreover, DNA is stable compared with other molecules and it can be stored for long periods of time.

DNA barcoding is an advanced and robust technique for species identification. The main advantage of using this technique is that DNA analysis can be done at any stage without the influence of plant organs and developmental stages. Along with fresh plant; dried, powdered or extract forms of plant material can also be used to extract DNA from original plants or processed products (Novak *et al.*, 2007). Once a DNA extraction has been completed only a small proportion of this is required for barcoding purposes and the rest can be stored for long periods of time (Howard, 2010). The samples can be identified by using short DNA sequences called DNA barcodes. The barcoding method is used to identify different species along with intraspecies relationships (Hebert, 2003). Nuclear and chloroplast regions are mainly used as plant DNA barcodes.

3.1.3 DNA test introduction into the British Pharmacopoeia

DNA Based Identification Techniques for Herbal Drugs were introduced in 2016 as a new Appendix method in the British Pharmacopoeia (British Pharmacopoeia Appendix XI V, 2017) (Sgamma *et al.*, 2017). This DNA based molecular identification is applicable to any investigation. DNA extraction, PCR, gel electrophoresis and DNA sequence analysis are general process steps of DNA barcoding method. The second section of supplementary chapter (British Pharmacopoeia Supplementary Chapter SC VII D, 2017) has illustrated the steps that are relevant to the monograph species. Along with providing a reference sequence for each subject, information is given out on requirement of DNA purification and usage of PCR primers.

The first case study of this identification method was *O. tenuiflorum*. At the first point, sample description, their analysis and how they are traded in the UK market, were

elaborated. This was based on industrial production processes and measurement of specific chemical formula for allopathic pharmaceutical drugs. As herbal drugs are assessed in their 'raw' form, their sample size and variety are more important. Later, during barcode region selection process, 5 regions were assessed to ascertain which was most efficient at distinguishing the commercial samples containing different *Ocimum* species. Among all these regions, *trnH-psbA* sequencing results could successfully discriminate among closely related *Ocimum* species. These results are also supported by published literature (Christina and Annamalai, 2014).

Holy basil (*O. tenuiflorum* L.) herbal drug material, used for BP project, were tested by DNA extraction, purification and amplification methods. Later, this reference *O. tenuiflorum* sequence was published which can be used to identify closely related species which contaminate Holy Basil herbal material. The same structure and method that was designed and implemented for Holy Basil material, can be used with different species as a model for high quality molecular characterisation. This data was generated by using DNA based identification method, has been merged along with the results of morphological and chemical analysis to ensure that only genuine details from good quality samples be used to produce the benchmark in the British Pharmacopoeia (Sgamma *et al.*, 2017).

3.1.4 Authentication of different *Ocimum* species

Different *Ocimum* plant species have different levels of eugenol, ME (methyl eugenol) compounds and therefore it is important to authenticate each species prior to initiate an investigation of these secondary metabolites. After authenticating Tulsi samples using DNA barcoding region, an identification of targeted secondary metabolites from different *Ocimum* species was initiated. Phenylpropanoids including eugenol, chavicol and their methyl derivatives are the major secondary metabolites within the *Ocimum* genus. There is very limited information available regarding relevant enzymes, pathway steps and how these secondary metabolites levels are modulated in different *Ocimum* species. Anand *et al.*, (2016) have studied different secondary metabolites levels within *O. basilicum*, *O. tenuiflorum* and *O. gratissimum*. Henceforth, after authenticating

unknown sample as *Ocimum*, certain phenylpropanoids (eugenol, ME, chavicol and estragole) identification was the next important and crucial step for this project.

3.1.5 High Performance Thin Layer Chromatography (HPTLC)

Analysis of new drugs, pharmaceuticals and natural compounds have required different tools in all the stages of drug discovery. High-performance thin layer chromatography (HPTLC) is one of these techniques which has widely been used in the pharmaceutical world. This technique is based on the principles of thin layer chromatography. HPTLC is an advanced form of TLC which provides more precise and accurate data. It also supports postmarket surveillance (Sweedler, 2002). The use of high-quality TLC plates and stationary phase accompanied with finer particle sizes allow better resolution (Reich and Schibli, 2005). The repeat development of the plate allows improved separation of results and a lower limit of detection.

The majority of modern pharmaceutical analysis techniques have expensive instruments and time-consuming applications. HPTLC has managed to overcome some of these problems. It has many advantages including minimum sample preparation, selective detection principle, automation and full optimization enable it to be an influential technique to get chromatographic details of natural products, pharmaceuticals and food stuff as well.

3.1.5.1 Chemovar identification through HPTLC

Natural variation occurs within plant species and can affect the phytochemical profile of the medicinal material. These differences can become stabilised, and subsequently result in the presence of ‘chemotypes’ within a plant species. Although these plants are from the same species, a ‘chemotype’ is a sub type within that species which represents particular chemical characteristics which can be identified via HPTLC methods. The characteristic profiles called ‘fingerprints’ are useful for identification as well as for quality control. For instance, chemtypes may result due to the different geographic locations in which a plant was grown. Traditional herbal medicine authenticity and quality can be checked through this chromatographic fingerprinting technique. Along with quality control, combination of other techniques including DNA authentication and

determining the identity of components in a mixture, can become an important tool in herbal drug manufacturing sectors (Choudhary and Sekhon, 2011). HPTLC has gained popularity as one of the ideal TLC techniques to study secondary metabolites levels in different plant species and chemotypes.

Two different cultivars of *O. tenuiflorum*, “Ram” and “Shyam” Tulsi are used in Ayurvedic, Sidhha and other traditional medicinal systems for treatment of malaria, bronchitis and other disorders. Their essential oils are reported as an insecticidal, antioxidant and antimicrobial agents (Pandey *et al.*, 2014). The composition of these essential oils varies due to genetic and environmental factors which influence the genetic expression of *Ocimum* plant.

HPTLC analysis of *Ocimum tenuiflorum* materials has indicated the production of either eugenol rich or ME rich essential oils (Grover and Rao, 1977). The HPTLC test was performed to investigate the relation between their essential oils and types of fingerprints. Both chemotypes were distinguished and Type 1 samples were identified as ME type essential oil, while Type 2 samples were identified as eugenol type essential oil (Figure 3-1) (Schinle *et al.*, 2014).

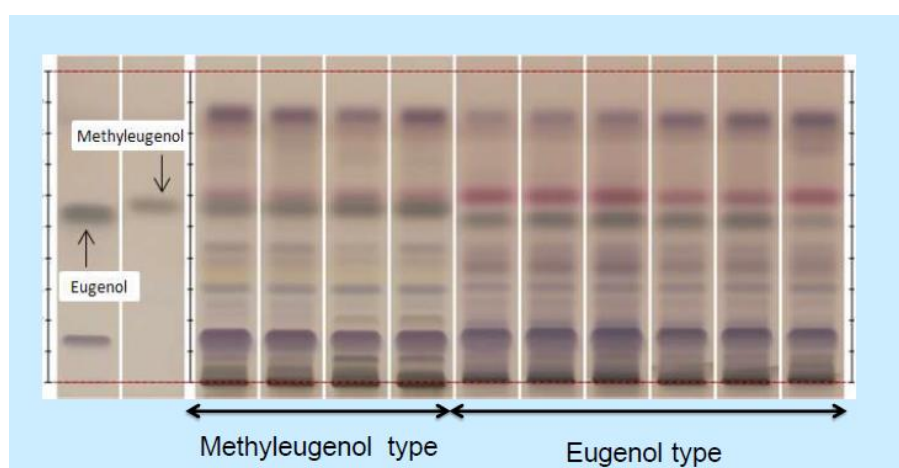


Figure 3-1: Essential oil fingerprint of 10 samples of Holy basil.
Adapted from Schinle *et al.*, 2014. The HPTLC plate differentiate between ME and eugenol type flavonoids in different *O. tenuiflorum* chemotypes.

3.1.6 DNA authentication of different *Ocimum* species

As different *Ocimum* species are morphologically very similar, they can be easily misidentified and could have harmful therapeutic effect while using them for medicinal purpose. To overcome this issue, DNA barcodes were used to validate some *Ocimum* species. Out of all (*matK*, *rbcL* and *trnH-psbA*) barcodes used to differentiate between *Ocimum* species, *trnH-psbA* was the best region to identify *O. tenuiflorum* species (Christina and Annamalai, 2014). In recent studies, *trnH-psbA* and *ITS* were used to differentiate between *O. tenuiflorum* and *O. gratissimum* among which *trnH-psbA* proved to be competent for PCR and sequence analysis (Bhamra, 2016).

3.1.7 Steps required for DNA barcoding

DNA extraction, PCR, checking PCR product on gel followed by sequencing and bioinformatics analysis are essential steps in DNA barcoding.

3.1.7.1 DNA extraction

The first stage of DNA extraction protocol is to disrupt the cell or nuclear membrane. Many commercial techniques suggest starting the extraction processes by crushing the plant material with liquid nitrogen to avoid DNA degradation. This macerating releases the DNA from the cell, and the cell walls and membranes are later collected in precipitation form along with the removal of primary metabolites including polysaccharides and proteins. Afterwards, DNA quality can be measured by using the ultraviolet absorption, diphenylamine reaction or fluorescence staining methods (Li, *et al.*, 2014). Readymade DNA extraction kits are flexible and can achieve high quality and yield of DNA compared to traditional DNA extraction techniques (Xin and Chen, 2006).

3.1.7.2 PCR of *trnH-psbA* region

PCR is a widely used molecular biology technique which can amplify the specific segment of DNA. This can generate thousands copies of selected DNA sequences. Denaturation, annealing and extension are the three main steps of the PCR process. The

trnH-psbA is the prominent region used throughout this project for authentication which successfully identified unknown *Ocimum* species. The PCR method has efficiently identified closely related species of *Ocimum* (*O. tenuiflorum*, *O. gratissimum*, *O. basilicum*). This technique is also useful to identify different species from the DNA samples mixture. Along with identifying different species, this technique can also be used to find adulterants. This is an extremely useful technique in industry sector for species level identification (Bhamra *et al.*, 2015).

3.1.7.3 Gel electrophoresis

Gel electrophoresis technique is used for isolation and separation of DNA fragments according to their size when current applied. As DNA carries a negative charge, fragments move towards the positive electrode when a current is applied to an agarose gel. Afterwards these separated, different size fragments are observed with the help of SYBR safe stain under UV rays.

The positive PCR samples, which were confirmed by gel electrophoresis were sent for sequencing.

3.1.7.4 Sequencing

DNA sequencing is the method to determine the position of four nucleotide bases (Adenine (A), guanine (G), cytosine (C), and thymine (T)) on DNA strands. Previously only ‘Sanger Sequencing’ was in practice which is still hugely popular, though ‘Next Generation Sequencing’ has been speedily overtaking this method. Sanger sequencing yields a limited amount of data while NGS can build up a very huge library out of just a single sequencing run.

3.1.7.4.1 Sanger sequencing

Fred Sanger and his colleagues developed this method which can sequence a fairly-long (around 900 bp) DNA fragment (Roe, 2014). Each of the fragments contained fluorescent labels which is the indication of sequencing termination processes. This technique is similar to the PCR reaction where multiple copies of a targeted DNA fragment are obtained, though it also contains dideoxy nucleotides which terminate the DNA

replication reaction. As each of these Dideoxy bases lack one hydroxyl group, as soon as they attached to the fragment no further bases can be added. By differently labelling dideoxy nucleotides with fluorophores relevant to their base (A, C, G or T), the signal can be interpreted to ascertain the base at each terminal residue. The reaction is optimised such that each base position is represented by a ‘terminated’ amplicon, each differing in length by one base. A long narrow gel matrix known as ‘Capillary Gel Electrophoresis’ separates these amplicons by size, and the fluorescent label attached to each is identified using a laser. These signals are then worked into an electropherogram, plotting elution time against fluorescent signal, which is then transformed into a base call at each residue.

3.1.7.4.2 *Contig preparation*

The process of assembling DNA sequences contains several steps including – trimming (process of removing ambiguous or poor quality sequence part), overlapping multiple DNA sequences to produce contigs (which represent a consensus region of DNA) and developing a consensus sequence using all of the data. It is reliable to use multiple and good quality reads to prepare contigs. Generally, forward and reverse reads are used to produce a reliable contig. Before using this consensus sequence as the reference DNA sequence, the conflicts need to be resolved. Taq polymerase enzyme used during PCR reaction, has low error rate. These errors are significantly important while working on large-scale cloning projects (Cline, 1996; McInerney *et al.*, 2014). By using more than one sequence read, this error rate can be accounted for allowing a reliable consensus sequence to be generated for a targeted region.

3.1.7.4.3 *Sequence alignment*

Alignment is one of the most commonly used tools in the bioinformatics world (Liu and Warnow, 2014). Any sequence data including DNA, AA, protein, genes, primers, promoters can be aligned together to find similarity and differences within sequences. This helps to understand structural, functional and evolutionary relationships of sequences included in alignment. Along with sequence analysis, alignment tool can also be used to study published, experiments and genomic data for comparison purposes.

3.1.7.4.4 *BLAST search*

The consensus sequence created by assembling multiple sequence reads was used to identify unknown samples collected throughout the investigation. The Basic Local Alignment Search Tool (BLAST) was used to investigate all of these sequences and to confirm the identity of the species. The species identification has been carried out on the basis of a 100% match to the reference sequence.

This chapter illustrates the use of DNA barcoding to authenticate *Ocimum* species. The materials and techniques used are described in next section.

3.2 Materials and methods

3.2.1 Plant Sample materials

There are a number of different types of plant samples used for this study which can be mainly divided into two categories (Table 3-1):

- Home grown materials - The project started with three different Tulsi samples, which had not had their botanical identity confirmed. Fresh plants, twigs and seeds of different *Ocimum* species were donated by members of the Hindu community from the UK, India and Egypt.
- Basil varieties from Greece – Dr Tiziana Sgamma found some local basil varieties (Table 3-1) during her trip to Greece. DNA was extracted from these for authentication purposes and their seeds were used for tissue culture process.

Table 3-1: Plant sample details including DNA extraction no., sample details and source information

DNA extraction No.	Sample details	Source details
1089	<i>Ram</i>	Home grown materials
1090	<i>Shaym</i>	
1091	<i>Vrinda</i>	
1314	BASIL - Ornamental 1	Sithonia, Greece
1315	BASIL - Ornamental 2	
1316	BASIL - Evergreen	
1317	BASIL - Dry specimens	
1318	BASIL - Baillicum 1	
1319	BASIL - Baillicum 2	

3.2.2 DNA samples

Apart from plant samples, some DNA vouchered specimens were used to identify *EOMT* region from previously authenticated *Ocimum* DNA samples.

- The CAMAG (with HPTLC information about the chemovars) – HPTLC test carried out by CAMAG (Schinle *et al.*, 2014) to identify Tulsi from other Lamiaceae family plants along with chemical differentiation of two chemotypes (green and black) of *O. tenuiflorum*. The two chemotypes are characterised by high levels of eugenol or methyl eugenol.
- Kew samples – The DNA samples, obtained from the Royal Botanic Gardens Kew - DNA bank, were also included in this study.
- BTG Bank samples – The previous BTG research student collected a wide range of *Ocimum* samples (Bhamra, 2016). These authenticated samples were used to identify *EOMT* gene for this project.
- Vouchered specimens – Some vouchered species specimens from Karlsruhe botanic garden were further analysed for this project.

Please find below details of these samples in Table 3-2:

Table 3-2: DNA sample details including sample no. sample reference no. (along with chemotypes details – Type1 or Type 2) and source information

Sample No.	Labels	Sample Ref. No.	Source
V86	<i>O. tenuiflorum</i>	S9890	CAMAG
		S9891 (Type 2)	
V87		S9892 (Type 2)	
		S9893 (Type 2)	
		S9888 (Type 2)	
V88		S10031 (Type 1)	
		S10032 (Type 1)	
		S10035	
V89		S10034 (Type 1)	
		S9891	
V90		S10038	
V91		S10042	
V92		S10527	
V93		S10528	
V94		S10529	
V95		S12200	
V96		S12201	
V97		S12202	
V98		S12203	
V99		S12609	
C81	<i>O. gratissimum</i>	1138	Kew Garden
C82	<i>O. selloi</i>	1583	
C84	<i>O. tenuiflorum</i>	9937	
C85	<i>O. americanum</i>	9939	
B05	<i>O. tenuiflorum</i>	437	BTG Bank samples
B13	<i>O. tenuiflorum</i>	461	
B15	<i>O. tenuiflorum</i>	491	
B16	<i>O. basilicum</i>	493	
G26	<i>O. tenuiflorum</i>	8256	Karlsruhe Institute of Technology, Germany
G27	<i>O. tenuiflorum</i>	8257	
G28	<i>O. africanum</i>	7537	
G29	<i>O. tenuiflorum</i>	5748	
G30	<i>O. gratissimum</i>	5749	
G31	<i>O. kilimandscharicum</i>	7810	
G32	<i>O. gratissimum</i>	8109	
G33	<i>O. tenuiflorum</i>	8258	
G34	<i>O. basilicum</i>	5192	
G35	<i>O. carnosum</i>	7564	
G36	<i>O. tenuiflorum</i>	5751	
G37	<i>O. tenuiflorum</i>	8097	
G38	<i>O. tenuiflorum</i>	8098	
G39	<i>O. tenuiflorum</i>	8099	

3.2.2.1 DNA samples - further processing

Some vouchered DNA samples obtained from London (CAMAG, NIBSC) and Germany (Prof. Dr. Peter Nick, the Molecular Cell Biology Unit, Karlsruhe Institute of Technology) were directly used for PCR with selected primers, though some of them

required clean-up/dilution to overcome lower concentration and contamination issues. (See Table 3-3 for more details).

Table 3-3: Biomolecular Technology Group *Ocimum* Samples

Species	Sample source	Sample type	Clean-up
<i>O. tenuiflorum</i>	Local community, Pukka, Germany, Kew and NIBSC	Fresh plant material, dry powder, vouchered DNA samples	58
<i>O. gratissimum</i>		Fresh plant material, vouchered DNA samples	2
<i>O. basillium</i>		Fresh plant material, vouchered DNA samples	1
<i>O. selloi</i>		Vouchered DNA samples	No clean up required
<i>O. americanum</i>		Vouchered DNA samples	2
<i>O. citriodorum</i>		Vouchered DNA samples	No clean up required
<i>O. kilimandscharicum</i>		Vouchered DNA samples	No clean up required
<i>O. carnosum</i>		Vouchered DNA samples	1

3.2.2.2 Sample collection

Fresh *Ocimum* plants were collected from the local Hindu community in Leicester. After detailed conversation with South Asian people, three different Tulsi varieties – ‘Ram’, ‘Shyam’ and ‘Vrinda’ were collected. They were later used for the first round of DNA extractions. There were a variety of samples collected for this project and their details are available in this chapter. As shown in figure 3-2, three different varieties of *Ocimum* were grown in the green house to check species variations and to compare them with a large collection of Tulsi DNAs at Biomolecular Technology Group’s database. Two different types of seeds (from Gujarat, India) and cutting (from Egypt) were donated by members of the Hindu community in Leicester. Two different seeds were used in this study- one called *Shyam* or *Krishna* Tulsi which is dark purplish in colour, while another one known as *Ram* Tulsi which is light green in colour. The third *Vrinda* variety has wide leaves in comparison to the other two. This one was grown from a cutting was more robust than the other plants.

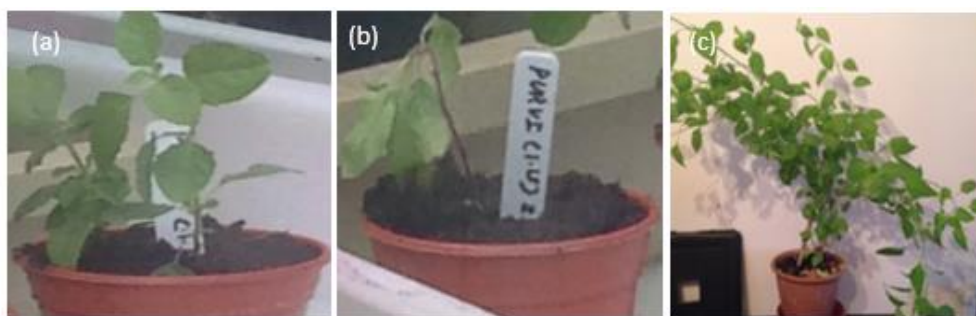


Figure 3-2: Three different *Ocimum* varieties (A) *Ram* (light green in colour) (B) *Shyam* or *Krishna* (dark purplish in colour) and (C) *Vrinda* variety (grows in natural surroundings without cultivation or care)

3.2.3 DNA isolation

DNA extractions were carried out according to Qiagen DNeasy Plant Mini Kit protocol for collected plant samples.

Fresh plant material was used for DNA extraction. 0.02 g dry plant sample (fresh samples kept aside for 1 week time) was mixed with 400 μ L Buffer AP1 and 4 μ L RNase A and a 3mm tungsten carbide bead. By using the TissueLyser Adapter set, this plant sample was crushed completely in two steps by reversing the position of tubes in each step for 1 min at 30 Hz.

The protocol instructions were followed from 3rd step of incubation at 65°C for 10 mins. Remaining steps were then followed to get the total volume of 200 μ L by doing two elution steps of 100 μ L.

3.2.3.1 *trnH-psbA* Barcode primers

This universal primer pair requires a specifically designed PCR reaction called ‘Touchdown Program’. For *trnH-psbA* primer pair, a range of annealing temperatures were set, starting from higher than primer’s T_m (the annealing temperature) temperature which was then gradually reduced by 1°C per cycle. Once the temperature was reached at the lowest level, normal PCR cycles were carried out.

trnH 5'-CGCGCATGGTGGATTCACAATCC-3'
psbA 5'-GTTATGCATGAACGTAATGCTC-3'

3.2.4 PCR to test DNA extractions and Plant barcoding

PCR reactions consisted of MyTaq Red Mix (2x), relevant primers (0.1µM each) and template DNA (0.4µL) made up to a final volume 10µL with nuclease free water in 0.2mL polypropylene tubes. All reactions were carried out in G-Storm GS1 Thermal Cycler and MJ Research PTC-200 Thermal Cycler. PCR programme details are included in Table 3-4.

Table 3-4: PCR programme for *trnH-psbA* primer pair

Primer	PCR programme
<i>trnH</i> <i>psbA</i>	5min at 95°C initial denaturation step, 35 cycles consisting of 1min at 95°C, 30s at touchdown temperature and 1min at 72°C, final extension period of 7min at 72°C. Touchdown temperature began at 58°C, reduced by 1°C per cycle until 48°C, then continued at 48°C for the remainder of the programme.

Thermocycling programme for *trnH-psbA* is different from other PCR programs (Fazekas *et al.*, 2008). A touchdown PCR cycling was used from 58°C to 48°C where temperature is reduced by 1°C at each step.

Reactions without DNA template were utilised as controls. PCR products were run on 2% (w/v) agarose gels in 0.5 X TBE buffer with 2µL SYBRsafe™ (Invitrogen, Carlsbad, CA, USA) DNA stain at 90V for 25 minutes and analysed with Gel Doc EZ system.

3.2.4.1 DNA clean up

The DNA clean-ups conducted on fresh *Ocimum* samples due to unsuccessful PCR amplification at first stage. An SOP (Standard Operating Procedure) steps of the isopropanol clean-up method can be seen in section 2.7. The isopropanol clean-up DNA samples were further diluted with double distilled water in 1:10 amount before conducting PCR.

3.2.5 Sample sequencing

The PCR samples which gave positive gel electrophoresis results, were sent for sequencing. The Sanger sequencing services provided by an external sequencing company called Macrogen (The Netherlands) was selected to obtain results. The result details contained sequence trace, GC% and the sequence length were later analysed in detail for identification purpose. These results were also used to construct a reference database.

3.2.5.1 Sequence analysis

The CLC main workbench was used to analyse the data where a series of processes including multiple DNA sequences assembly, trimming, overlapping sequences to produce consensus sequence (contig preparation) and alignment were carried out to determine the species identity. Further details about sequence analysis could be earlier in this chapter.

The BLAST (The Basic Local Alignment Search Tool) search tool (<http://blast.ncbi.nlm.nih.gov>) was used to compare the query DNA sequences with reference DNA barcodes included in databases, for identification purpose. BOLD (<http://www.boldsystems.org>) and Genbank (<http://www.ncbi.nlm.nih.gov/genbank/>) were used to get reference *Ocimum* sequences. These reference sequences were later compared with the query sequences in CLC software.

3.3 Result

3.3.1 Gel electrophoresis: *trnH-psbA* region

Home grown Tulsi samples which were mainly characterised in three groups – *Ram* (DNA extraction no. - 1089), *Shyam* (DNA extraction no. – 1090) and *Vrinda* (DNA extraction no. – 1091) were further examined by using *trnH-psbA* primer pair. As shown in figure 3-3, DNA samples were used in different forms – normal (Lane 1, 2 and 3), isopropanol washed (Lane 4, 5 and 6) and diluted (1:10) samples (Lane 7, 8 and 9) where isopropanol washed samples work better than remaining samples.

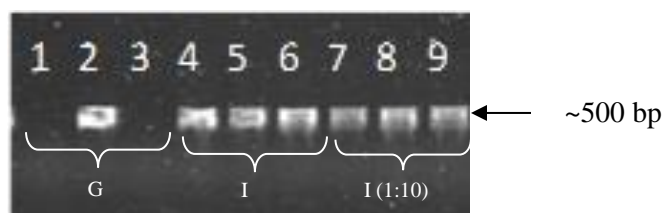


Figure 3-3: *trnH-psbA* PCR of *Ram* sample (1, 4 and 7), *Shyam* sample (2, 5 and 8) and *Vrinda* sample (3, 6 and 9). Genomic (G) DNA extractions – Lane 1, *Ram* sample; Lane 2, *Shyam* sample and Lane 3, *Vrinda* sample. Isopropanol (I) washed DNA extractions – Lane 4, *Ram* sample; Lane 5, *Shyam* sample and Lane 6, *Vrinda* sample. Isopropanol washed and diluted [I (1:10)] extractions – Lane 7, *Ram* sample; Lane 8, *Shyam* sample and Lane 9, *Vrinda* sample.

Basil samples collected from Greece (Table 3-1), were also authenticated by using *trnH-psbA* primer pair where relevant region presence was confirmed through PCR reaction as shown in figure 3-4. The band can be seen the positive control and lanes 1 (DNA extraction no. – 1314), lane 2 (DNA extraction no. – 1315), lane 3 (DNA extraction no. – 1316), lane 5 (where more than one band present) (DNA extraction no. – 1318) and lane 6 (DNA extraction no. – 1319); while no band was present in lane 4 (DNA extraction no. – 1317) and the negative control. The expected product sizes (~550 bp) are present in the gel picture.

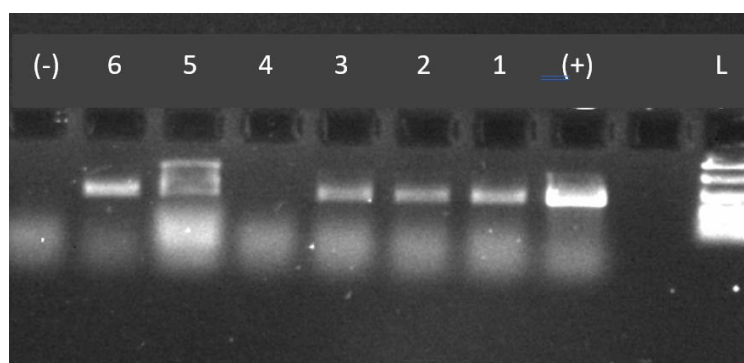


Figure 3-4: *trnH-psbA* PCR for Basil DNA extractions. (+), positive control; Lane 1, BASIL - Ornamental 1 sample; Lane 2, BASIL - Ornamental 2; Lane 3, BASIL – Evergreen; Lane 4, BASIL - Dry specimens; Lane 5, BASIL - Basillicum 1; Lane 6, BASIL - Basillicum 2, (-), negative control.

3.3.2 DNA authentication of unknown Tulsi samples

The following sections contain the details regarding how unknown Tulsi samples were analysed using *trnH-psbA* barcoding region.

3.3.2.1 Sequence analysis of *trnH-psbA* region

Out of several collected samples, the majority were identified as *Ram* or *Shyam* types while very few were identified as *Vrinda* Tulsi type. The successfully amplified *trnH-psbA* sequences were used to prepare contigs and consensus sequences after resolving conflicts (Figure 3-5).

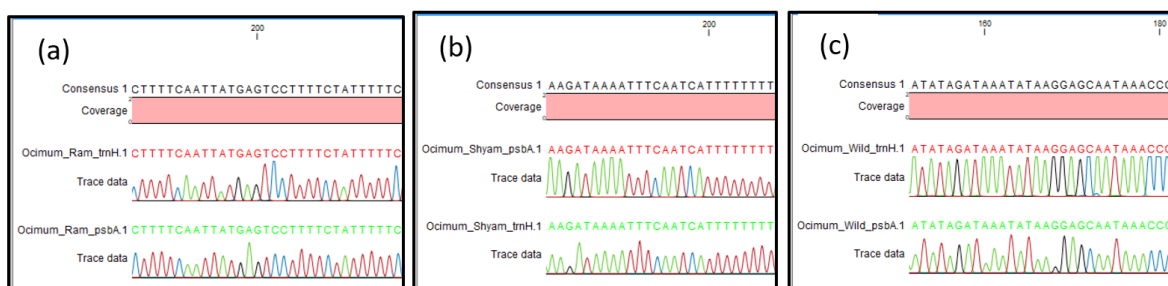


Figure 3-5: Electropherograms of *trnH-psbA* sequences of three Tulsi types. (a) *Ram* Tulsi (b) *Shyam* Tulsi and (c) *Vrinda* Tulsi.

The consensus sequence was prepared by assembling both (*trnH-psbA*) reads and were used to identify the unknown Tulsi samples collected throughout the investigation. The identification of the samples was confirmed by entering and comparing their sequences into BLAST with other sequences in the database.

3.3.2.2 The BLAST search

A BLAST search (<http://blast.ncbi.nlm.nih.gov/>) was conducted on the sequence to observe the closeness of the sequence identity and similarity from the NCBI database.

As shown in Figure 3-6, *Ram* and *Shyam* consensus sequences were very similar to *O. tenuiflorum* sequences, while *Vrinda* consensus sequence was similar to *O. gratissimum* reference sequences. To confirm their identity, these samples sequences were further aligned with the known reference sequences in the database.

(a)	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Ocimum tenuiflorum voucher CIMAP:60 photosystem II protein D1 (psbA) gene, partial cds; and tRNA-His (trnH) gene, partial sequence; chloroplast	752	752	91%	0.0	99%	MF457806.1
<input type="checkbox"/>	Ocimum tenuiflorum voucher CIMAP:64 photosystem II protein D1 (psbA) gene, partial cds; and tRNA-His (trnH) gene, partial sequence; chloroplast	752	752	91%	0.0	99%	MF457805.1
<input type="checkbox"/>	Ocimum tenuiflorum voucher CIMAP:68 photosystem II protein D1 (psbA) gene, partial cds; and tRNA-His (trnH) gene, partial sequence; chloroplast	752	752	91%	0.0	99%	MF457804.1
<input type="checkbox"/>	Ocimum tenuiflorum voucher CIMAP:63 photosystem II protein D1 (psbA) gene, partial cds; and tRNA-His (trnH) gene, partial sequence; chloroplast	752	752	91%	0.0	99%	MF457803.1
<input type="checkbox"/>	Ocimum tenuiflorum voucher CIMAP:61 photosystem II protein D1 (psbA) gene, partial cds; and tRNA-His (trnH) gene, partial sequence; chloroplast	752	752	91%	0.0	99%	MF457801.1

(b)	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Ocimum tenuiflorum voucher CIMAP:60 photosystem II protein D1 (psbA) gene, partial cds; and tRNA-His (trnH) gene, partial sequence; chloroplast	741	741	87%	0.0	99%	MF457806.1
<input type="checkbox"/>	Ocimum tenuiflorum voucher CIMAP:64 photosystem II protein D1 (psbA) gene, partial cds; and tRNA-His (trnH) gene, partial sequence; chloroplast	741	741	87%	0.0	99%	MF457805.1
<input type="checkbox"/>	Ocimum tenuiflorum voucher CIMAP:68 photosystem II protein D1 (psbA) gene, partial cds; and tRNA-His (trnH) gene, partial sequence; chloroplast	741	741	87%	0.0	99%	MF457804.1
<input type="checkbox"/>	Ocimum tenuiflorum voucher CIMAP:63 photosystem II protein D1 (psbA) gene, partial cds; and tRNA-His (trnH) gene, partial sequence; chloroplast	741	741	87%	0.0	99%	MF457803.1

(c)	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Ocimum gratissimum voucher CIMAP:47 photosystem II protein D1 (psbA) gene, partial cds; and tRNA-His (trnH) gene, partial sequence; chloroplast	719	719	76%	0.0	100%	MF457820.1
<input type="checkbox"/>	Ocimum gratissimum voucher CIMAP:48 photosystem II protein D1 (psbA) gene, partial cds; and tRNA-His (trnH) gene, partial sequence; chloroplast	713	713	76%	0.0	99%	MF457818.1
<input type="checkbox"/>	Ocimum gratissimum voucher CIMAP:45 photosystem II protein D1 (psbA) gene, partial cds; and tRNA-His (trnH) gene, partial sequence; chloroplast	706	706	76%	0.0	99%	MF457819.1
<input type="checkbox"/>	Ocimum gratissimum chloroplast trnH-psbA intergenic spacer, specimen voucher MIB:zpl:02997	697	697	74%	0.0	100%	FR726109.1
<input type="checkbox"/>	Ocimum gratissimum trnH-psbA intergenic spacer, partial sequence; chloroplast	676	676	72%	0.0	100%	KR735719.1

Figure 3-6: BLAST search results showed percentage similarity of *Ram*, *Shyam* and *Vrinda* samples. (a) and (b) *Ram* consensus sequences were showing highest similarities towards *O. tenuiflorum* sequences available in database. (c) *Vrinda* consensus sequence was showing similarity towards *O. gratissimum* sequences available in the database.

3.3.2.3 Reference barcode sequences

The reliable reference barcodes obtained from the GenBank and NCBI databases, were used to further identify the Tulsi samples. The reference sequences designed throughout this study will also be submitted to the databases.

3.3.2.3.1 *O. gratissimum* and *O. tenuiflorum* Reference Specimen

The previously recognised *O. gratissimum* sequences (Bhamra, 2016) and C81 sample sequence from The Royal Botanic Gardens Kew - DNA bank (Chase, *et al.*, 2005), were used as reference sequences to compare ‘*Vrinda*’ plant sequences. As all these samples have been identified and submitted to database by previous researchers as *O. gratissimum*, the experiment data sequences were aligned in a multiple alignment (Figure 3-7) to confirm their identity.

Similarly, a standard *O. tenuiflorum* sequence (Figure 3-8) was used to compare *Ram* and *Shyam* plant sequences. Earlier, they were also cross checked through BLAST search tool to check their closely related sequences from database.

Eight *O. gratissimum* *trnH-psbA* sequences (433, 437, 507, 490, 491, 769 and 1138) and the standard *O. tenuiflorum* sequence (which is published in British pharmacopoeia) were used for multiple sequence alignment (as shown in Table 3-5):

Table 3-5: Biomolecular Technology Group sequences.

Sr. No.	Species ID	Reference
1	<i>O. gratissimum</i> 433	B02 (Bhamra, 2016)
2	<i>O. gratissimum</i> 437	B05 (Bhamra, 2016)
3	<i>O. gratissimum</i> 507	B07 (Bhamra, 2016)
4	<i>O. gratissimum</i> 490	A14 (Bhamra, 2016)
5	<i>O. gratissimum</i> 491	B15 (Bhamra, 2016)
6	<i>O. gratissimum</i> 768	G30) (Accession No. - MF784560) Germany, Peter Nick Sample)
7	<i>O. gratissimum</i> 769	G32 (Germany, Peter Nick Sample)
8	<i>O. gratissimum</i> 1138	C81 (Accession No. - MF784559) (Kew DNA Bank)
9	Standard <i>O. tenuiflorum</i> sequence	Published in BP

O. gratissimum reference specimens were obtained from previous BTG research student (Bhamra, 2016) collection (433, 437, 507, 490 and 491), Botanical Institute Karlsruhe Institute of Technology, Germany (768 and 769) and The Royal Botanic Gardens Kew (1138) (Chase *et al.*, 2005). These previously identified specimens of *O. gratissimum* were used as reference sequences in the analysis. In alignment (Figure 3-7), the highlighted box (across 293 to 297 bases) shows some differences within these sequences. On the basis of their differences, they could be categorised in three types of *O. gratissimum* sequences. Sample No. 433, 437 and 507 were included in first group; Sample No. 490, 768 and 491 were included in second group along with ‘*Vrinda*’ sample sequence; 769 and 1138 were included in the third group.

```

1                                     50
433|      AT AACTTCCCTC TAGACCTAGC TGCTATTGAA GCTCCAACAA
437|      AT AACTTCCCTC TAGACCTAGC TGCTATTGAA GCTCCAACAA
507|      AT AACTTCCCTC TAGACCTAGC TGCTATTGAA GCTCCAACAA
490|      AT AACTTCCCTC TAGACCTAGC TGCTATTGAA GCTCCAACAA
Vrinda GAATGCTCAT AACTTCCCTC TAGACCTAGC TGCTATTGAA GCTCCAACAA
768|      AT AACTTCCCTC TAGACCTAGC TGCTATTGAA GCTCCAACAA
491|      AT AACTTCCCTC TAGACCTAGC TGCTATTGAA GCTCCAACAA
769|      AT AACTTCCCTC TAGACCTAGC TGCTATTGAA GCTCCAACAA
1138|      TAGACCTAGC TGCTATTGAA GCTCCAACAA
Consensus .....at aacttccctc TAGACCTAGC TGCTATTGAA GCTCCAACAA

51                                     100
433|  ATGGCTAAGA CTTGTTCTTG GTGTGTAGAA GTTTTTGAAA TATAGATAAA
437|  ATGGCTAAGA CTTGTTCTTG GTGTGTAGAA GTTTTTGAAA TATAGATAAA
507|  ATGGCTAAGA CTTGTTCTTG GTGTGTAGAA GTTTTTGAAA TATAGATAAA

```

490| ATGGCTAAGA CTTGTTCTTG GTGTGTAGAA GTTTTTGAAA TATAGATAAA
Vrinda ATGGCTAAGA CTTGTTCTTG GTGTGTAGAA GTTTTTGAAA TATAGATAAA
768| ATGGCTAAGA CTTGTTCTTG GTGTGTAGAA GTTTTTGAAA TATAGATAAA
491| ATGGCTAAGA CTTGTTCTTG GTGTGTAGAA GTTTTTGAAA TATAGATAAA
769| ATGGCTAAGA CTTGTTCTTG GTGTGTAGAA GTTTTTGAAA TATAGATAAA
1138| ATGGCTAAGA CTTGTTCTTG GTGTGTAGAA GTTTTTGAAA TATAGATAAA
Consensus ATGGCTAAGA CTTGTTCTTG GTGTGTAGAA GTTTTTGAAA TATAGATAAA

101 150
433| TATAAGGAGC AATAAACCCCT TTCTTGTTCT ATCAAAAGAG GGTTTATTGC
437| TATAAGGAGC AATAAACCCCT TTCTTGTTCT ATCAAAAGAG GGTTTATTGC
507| TATAAGGAGC AATAAACCCCT TTCTTGTTCT ATCAAAAGAG GGTTTATTGC
490| TATAAGGAGC AATAAACCCCT TTCTTGTTCT ATCAAAAGAG GGTTTATTGC
Vrinda TATAAGGAGC AATAAACCCCT TTCTTGTTCT ATCAAAAGAG GGTTTATTGC
768| TATAAGGAGC AATAAACCCCT TTCTTGTTCT ATCAAAAGAG GGTTTATTGC
491| TATAAGGAGC AATAAACCCCT TTCTTGTTCT ATCAAAAGAG GGTTTATTGC
769| TATAAGGAGC AATAAACCCCT TTCTTGTTCT ATCAAAAGAG GGTTTATTGC
1138| TATAAGGAGC AATAAACCCCT TTCTTGTTCT ATCAAAAGAG GGTTTATTGC
Consensus TATAAGGAGC AATAAACCCCT TTCTTGTTCT ATCAAAAGAG GGTTTATTGC

151 200
433| TCCTTTCTTT TCTTTTCAAT TAGGAGTCCT TTTCTATTTT TCTAGTAGTA
437| TCCTTTCTTT TCTTTTCAAT TAGGAGTCCT TTTCTATTTT TCTAGTAGTA
507| TCCTTTCTTT TCTTTTCAAT TAGGAGTCCT TTTCTATTTT TCTAGTAGTA
490| TCCTTTCTTT TCTTTTCAAT TAGGAGTCCT TTTCTATTTT TCTAGTAGTA
Vrinda TCCTTTCTTT TCTTTTCAAT TAGGAGTCCT TTTCTATTTT TCTAGTAGTA
768| TCCTTTCTTT TCTTTTCAAT TAGGAGTCCT TTTCTATTTT TCTAGTAGTA
491| TCCTTTCTTT TCTTTTCAAT TAGGAGTCCT TTTCTATTTT TCTAGTAGTA
769| TCCTTTCTTT TCTTTTCAAT TAGGAGTCCT TTTCTATTTT TCTAGTAGTA
1138| TCCTTTCTTT TCTTTTCAAT TAGGAGTCCT TTTCTATTTT TCTAGTAGTA
Consensus TCCTTTCTTT TCTTTTCAAT TAGGAGTCCT TTTCTATTTT TCTAGTAGTA

201 250
433| TTGGACTTAC CTAGACTTTT CTTTCCATAA AAGCAAGAAG ATAAAAAAAAAT
437| TTGGACTTAC CTAGACTTTT CTTTCCATAA AAGCAAGAAG ATAAAAAAAAAT
507| TTGGACTTAC CTAGACTTTT CTTTCCATAA AAGCAAGAAG ATAAAAAAAAAT
490| TTGGACTTAC CTAGACTTTT CTTTCCATAA AAGCAAGAAG ATAAAAAAAAAT
Vrinda TTGGACTTAC CTAGACTTTT CTTTCCATAA AAGCAAGAAG ATAAAAAAAAAT
768| TTGGACTTAC CTAGACTTTT CTTTCCATAA AAGCAAGAAG ATAAAAAAAAAT
491| TTGGACTTAC CTAGACTTTT CTTTCCATAA AAGCAAGAAG ATAAAAAAAAAT
769| TTGGACTTAC CTAGACTTTT CTTTCCATAA AAGCAAGAAG ATAAAAAAAAAT
1138| TTGGACTTAC CTAGACTTTT CTTTCCATAA AAGCAAGAAG ATAAAAAAAAAT
Consensus TTGGACTTAC CTAGACTTTT CTTTCCATAA AAGCAAGAAG ATAAAAAAAAAT

251 300
433| GATTGAAATT CTATCTTTTG TTTTACAATT TTGAAAAAA AAATTT-AAA
437| GATTGAAATT CTATCTTTTG TTTTACAATT TTGAAAAAA AAATTT-AAA
507| GATTGAAATT CTATCTTTTG TTTTACAATT TTGAAAAAA AAATTT-AAA
490| GATTGAAATT CTATCTTTTG TTTTACAATT TTGAAAAAA AA-TTT-AAA
Vrinda GATTGAAATT CTATCTTTTG TTTTACAATT TTGAAAAAA AA-TTT-AAA
768| GATTGAAATT CTATCTTTTG TTTTACAATT TTGAAAAAA AA-TTT-AAA
491| GATTGAAATT CTATCTTTTG TTTTACAATT TTGAAAAAA AA-TTT-AAA
769| GATTGAAATT CTATCTTTTG TTTTACAATT TTGAAAAAA AAATTC-AAA
1138| GATTGAAATT CTATCTTTTG TTTTACAATT TTGAAAAAA AAATTC-AAA
Consensus GATTGAAATT CTATCTTTTG TTTTACAATT TTGAAAAAA AaattT. AAA

301 350
433| TTGAAAAAGT GAATTCGTAA ATAAATAGAA AATTGACTTT TTTATAGTAG
437| TTGAAAAAGT GAATTCGTAA ATAAATAGAA AATTGACTTT TTTATAGTAG
507| TTGAAAAAGT GAATTCGTAA ATAAATAGAA AATTGACTTT TTTATAGTAG
490| TTGAAAAAGT GAATTCGTAA ATAAATAGAA AATTGACTTT TTTATAGTAG
Vrinda TTGAAAAAGT GAATTCGTAA ATAAATAGAA AATTGACTTT TTTATAGTAG
768| TTGAAAAAGT GAATTCGTAA ATAAATAGAA AATTGACTTT TTTATAGTAG
491| TTGAAAAAGT GAATTCGTAA ATAAATAGAA AATTGACTTT TTTATAGTAG
769| TTGAAAAAGT GAATTCGTAA ATAAATAGAA AATTGACTTT TTTATAGTAG

```

1138| TTGAAAAAGT GAATTCGTAA ATAAATAGAA AATTGACTTT TTTATAGTAG
Consensus TTGAAAAAGT GAATTCGTAA ATAAATAGAA AATTGACTTT TTTATAGTAG

351 392
433| AGGGGCGGAT GTAGCCAAGT GGATCAAG
437| AGGGGCGGAT GTAGCCAAGT GGATCAAG
507| AGGGGCGGAT GTAGCCAAGT GGATCAAG
490| AGGGGCGGAT GTAGCCAAGT GGATCAAG
Vrinda AGGGGCGGAT GTAGCCAAGT GGATCAAGGC AGTGGATTGT GA
768| AGGGGCGGAT GTAGCCAAGT GGATCAAG
491| AGGGGCGGAT GTAGCCAAGT GGATCAAG
769| AGGGGCGGAT GTAGCCAAGT GGATCA
1138| AGGGGCGGAT GTAGCCAAGT GGATCAAG
Consensus AGGGGCGGAT GTAGCCAAGT GGATCAag.. . . . . . . . . . .

```

Figure 3-7: Multiple sequence alignment of several *O. gratissimum* sequences with ‘Vrinda’ Tulsi sequences. The highlighted box shows three types of *O. gratissimum* sequences. Sample No. 433, 437 and 507 were included in first group; Sample No. 490, 768 and 491 were included in second group along with ‘Vrinda’ sample sequence; 769 and 1138 were included in third group.

According to BLAST search, it was found that the standard sequences of *O. gratissimum* were closely related to the ‘Vrinda’ Tulsi sample. Apart from highlighted part, experiment set of sequence had a perfect match with the reference sequences seen within the *trnH-psbA* region. Along with confirming query sequence similarity through BLAST search, aligning it with reference sequences helped to produce reliable results. The standard *O. tenuiflorum* sequence from *Ocimum* genus was published in BP. This sequence was aligned with *Ram* and *Shyam* sample sequences (as they were identified as *O. tenuiflorum* during BLAST search). As shown in Figure 3-8, the sequences were all very similar which confirms query sequences identity as *O. tenuiflorum*.

```

1 50
Standard_T GTTATGCATG AACGTAATGC TCATAACTTC CCTCTAGACC TAGCTGCTAT
Ram_trnH_p ACGTAATGC TCATAACTTC CCTCTAGACC TAGCTGCTAT
Shyam_trnH GTAATGC TCATAACTTC CCTCTAGACC TAGCTGCTAT
Consensus .....acGTAATGC TCATAACTTC CCTCTAGACC TAGCTGCTAT

51 100
Standard_T TGAAGCTCCA ACAAATGGCT AAGACTTGTT CTTGGTGTGT AGAAGTTTTT
Ram_trnH_p TGAAGCTCCA ACAAATGGCT AAGACTTGTT CTTGGTGTGT AGAAGTTTTT
Shyam_trnH TGAAGCTCCA ACAAATGGCT AAGACTTGTT CTTGGTGTGT AGAAGTTTTT
Consensus TGAAGCTCCA ACAAATGGCT AAGACTTGTT CTTGGTGTGT AGAAGTTTTT

101 150
Standard_T GAAATATAGA TAAATATAAG GAGCAATAAA CCCTTTCTTG TTCTATCAAA
Ram_trnH_p GAAATATAGA TAAATATAAG GAGCAATAAA CCCTTTCTTG TTCTATCAAA
Shyam_trnH GAAATATAGA TAAATATAAG GAGCAATAAA CCCTTTCTTG TTCTATCAAA
Consensus GAAATATAGA TAAATATAAG GAGCAATAAA CCCTTTCTTG TTCTATCAAA

151 200
Standard_T AGAGGGTTTA TTGCTCCTTT CTTTCTTTT CAATTATGAG TCCTTTTCTA
Ram_trnH_p AGAGGGTTTA TTGCTCCTTT CTTTCTTTT CAATTATGAG TCCTTTTCTA
Shyam_trnH AGAGGGTTTA TTGCTCCTTT CTTTCTTTT CAATTATGAG TCCTTTTCTA
Consensus AGAGGGTTTA TTGCTCCTTT CTTTCTTTT CAATTATGAG TCCTTTTCTA

201 250
Standard_T TTTTCTAGT AGTATTGGAC TTACCTAGAC TTTTCTTTCC ATAAAAGAAA
Ram_trnH_p TTTTCTAGT AGTATTGGAC TTACCTAGAC TTTTCTTTCC ATAAAAGAAA

```

<i>Shyam_trnH</i>	TTTTTCTAGT	AGTATTGGAC	TTACCTAGAC	TTTTCTTTCC	ATAAAAGAAA
Consensus	TTTTTCTAGT	AGTATTGGAC	TTACCTAGAC	TTTTCTTTCC	ATAAAAGAAA

	251				300
Standard_T	GAAGATAAAA	AAAATGATTG	AAATTTTATC	TTTTGTTTTA	CAATTTTGCA
<i>Ram_trnH_p</i>	GAAGATAAAA	AAAATGATTG	AAATTTTATC	TTTTGTTTTA	CAATTTTGCA
<i>Shyam_trnH</i>	GAAGATAAAA	AAAATGATTG	AAATTTTATC	TTTTGTTTTA	CAATTTTGCA
Consensus	GAAGATAAAA	AAAATGATTG	AAATTTTATC	TTTTGTTTTA	CAATTTTGCA

	301				350
Standard_T	AAAAAAATTC	AAATTGAAAA	AGTGAATTCG	TAAATAAATA	GAAAATTTTC
<i>Ram_trnH_p</i>	AAAAAAATTC	AAATTGAAAA	AGTGAATTCG	TAAATAAATA	GAAAATTTTC
<i>Shyam_trnH</i>	AAAAAAATTC	AAATTGAAAA	AGTGAATTCG	TAAATAAATA	GAAAATTTTC
Consensus	AAAAAAATTC	AAATTGAAAA	AGTGAATTCG	TAAATAAATA	GAAAATTTTC

	351				400
Standard_T	TTTTTAGATT	TTTAGAGTAG	AGGGGCGGAT	GAGCCAAGT	GGATCAAGGC
<i>Ram_trnH_p</i>	TTTTTAGATT	TTTAGAGTAG	AGGGGCGGAT	GAGCCAAGT	GGATCAAGGC
<i>Shyam_trnH</i>	TTTTTAGATT	TTTAGAGTAG	AGGGGCGGAT	GAGCCAAGT	GGATCAAGGC
Consensus	TTTTTAGATT	TTTAGAGTAG	AGGGGCGGAT	GAGCCAAGT	GGATCAAGGC

	401			427	
Standard_T	AGTGGATTGT	GAATCCACCC	ATGCGCG		
<i>Ram_trnH_p</i>	AGTGGATTGT	GAACCCC			
<i>Shyam_trnH</i>	AGTGGATTGT	GATCACCC			
Consensus	AGTGGATTGT	GAaccCcc..		

Figure 3-8: Multiple sequence alignment including standard sequence of *O. tenuiflorum* with 'Ram' and 'Shyam' type Tulsi sequences. The standard *O. tenuiflorum* trnH-psbA sequence was published in British Pharmacopoeia Appendix XI V, 2017 used as reference sequence to identify 'Ram' and 'Shyam' Tulsi samples.

3.4 Discussion

3.4.1 trnH-psbA

Compared to the nrITS region, the *trnH-psbA* region proved to be more useful for species level identification (Bhamra, 2016). An identification through *trnH-psbA* region is an effective tool to use within *Ocimum* genus. This region is amplified, aligned and tested with published *Ocimum* sequences for DNA based assays. The divergence within different *Ocimum* species created authentication issues while designing PCRs for some species, as targeted amplicon was either too long or too short. Though with the help of isopropanol clean up and dilution steps, all collected *Ocimum* samples were successfully identified using *trnH-psbA* barcode region.

An interspecific variation amongst *trnH-psbA* region show high variation amongst genus *Ocimum*. The result output and sequence quality were good for this amplified region. Different *Ocimum* species along with refence sequence were aligned together to

identify unknown species. All results were personally analysed by doing DNA extraction from collected samples. A number of samples submitted to the general DNA extraction method did not yield amplifiable DNA, and therefore clean-up and dilutions were carried out. This helps to remove inhibitors either from the sample or introduced during DNA extraction process and succeeded in producing good quality results. In short, the quality of the results was sufficient in the first round of experiments following the method described in the British pharmacopoeia (2016), *trnH-psbA* region can be successfully used to authenticate *Ocimum* species.

3.4.1.1 Universal barcode primers

The strategy of using relevant primer pairs covering *trnH-psbA* barcoding region, has been successfully implemented during this project. The high level of interspecific variation within the *Ocimum* genus has been successfully applied by studying the *trnH-psbA* barcoding region. The region amplified using PCR technique was helpful to get good quality sequence data. The *trnH-psbA* region used to discriminate between different *Ocimum* species in this project. PCR test with *trnH-psbA* primer pair followed by BLAST search results, used to identify variety of *Ocimum* samples (fresh and dry plant material along with commercial DNA samples). The final results have appeared to work more efficiently with clean-up and diluted DNA samples as this helped to remove PCR inhibitors. Eventually, the good quality sequence data has been acquired from most of the samples using bioinformatics tools (Please see section 3.2 for more details).

3.4.2 Advantage of using DNA based technique

Barcode primers covering short bp sequence (less than 500 bp for *trnH-psbA* region) gave better results compared to primers covering long bp sequences (around 1000 bp for nrITS region) (Bhamra, 2016). This was further confirmed by agarose gel electrophoresis results where a bright band can be seen, indicating a large product copy number. Within this project a product of around 200-300 bp is suitable for use of all available DNA samples. Morphological identification is not sufficient to identify processed plant material

and the different chemotypes shown in the HPTLC analysis are a prime example of the difficulties in using these methods for species identification. Thus, DNA based identification method is the best method for species identification purpose.

3.4.3 In detail investigation

DNA extraction was initiated by collecting different *Ocimum* species. The BLAST search for most of these samples highlighted similarity with relevant *Ocimum* species. The previous literature revealed that misidentification of *Ocimum* species for drug manufacturing purpose was the main issue which can be overcome by using DNA based techniques along with using results achieved during this research as a reference. Although, it can be difficult to obtain good quality result data while using commercial products including tablets, capsules, extracts and tincture (Jürges *et al.*, 2009). This chapter shows the possibility of using DNA barcode region for identification purposes, when other techniques failed to provide positive results. This technique is accessible and can be used in all sectors including industry, pharmaceutical and research environment.

3.4.4 Contig description

When a good quality DNA sequence was obtained, two reads (forward and reverse) were ordered to create reliable contig (is a set of overlapping DNA segments that together represent a consensus region of DNA) (Figure 3-9). All conflicts were resolved before consensus sequence used as the reference DNA sequence; to compare with available sequences in database through BLAST search and prepare a multiple alignment.

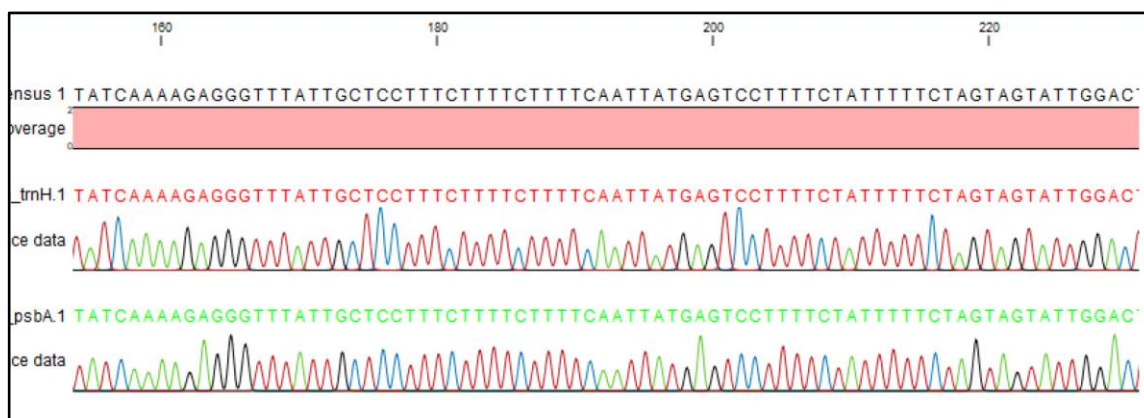


Figure 3-9: A Section of a Contig Produced from two Electropherograms of *trnH-psbA* Sequences from query sample

As shown in figure 3-5, the sequencing results were of high quality for all three Tulsi (*Ram*, *Shyam* and *Vrinda*) types. Based on these results, reliable contigs and consensus sequences were produced. Several reference *Ocimum* DNA sequences obtained from databases, were used to identify unknown *Ocimum* species including *O. tenuiflorum*, *O. gratissimum*, *O. africanum*, *O. americanum*, and *O. citriodorum*, *O. basilicum*, *O. kilimandscharicum* and *O. selloi*.

3.4.5 Authentication of unknown samples

The *trnH-psbA* sequences results confirmed the selected specimens as *Ocimum*. The identity of all of these sequences was confirmed through BLAST search in NCBI database along with aligning them with reference sample sequences. Table 3-6 summarises the sequence data which were created to identify *trnH-psbA* region.

Table 3-6: Samples Identification Summary

Sample ID:	Sample source:	Type of tulsi specified:	<i>trnH-psbA</i> sequence
G26	Dr. Willmar Schwabe Pharmaceuticals, Germany	<i>O. tenuiflorum</i>	✓
G27		<i>O. tenuiflorum</i>	✓
G28		<i>O. africanum</i>	✓
G29		<i>O. citridorum</i>	✓
G30		<i>O. gratissimum</i>	✓
G31		<i>O. kilimandscharicum</i>	✓
G32		<i>O. gratissimum</i>	✓
G33		<i>O. tenuiflorum</i>	✓
G34		<i>O. basilicum</i>	✓
G36		<i>O. tenuiflorum</i>	✓
G37		<i>O. tenuiflorum</i>	✓
G38		<i>O. tenuiflorum</i>	✓
G39		<i>O. tenuiflorum</i>	✓
C81	Kew	<i>O. gratissimum</i>	✓
C82		<i>O. selloi</i>	X
C84		<i>O. tenuiflorum</i>	✓
C85		<i>O. americanum</i>	✓
V86	CAMAG	<i>O. tenuiflorum</i>	✓
V87		<i>O. tenuiflorum</i>	✓
V88		<i>O. tenuiflorum</i>	✓
V90		<i>O. tenuiflorum</i>	✓
V92		<i>O. tenuiflorum</i>	✓
V93		<i>O. tenuiflorum</i>	✓
V94		<i>O. tenuiflorum</i>	✓
V96		<i>O. tenuiflorum</i>	✓
V98		<i>O. tenuiflorum</i>	✓
V99		<i>O. tenuiflorum</i>	✓
B05	BTG	<i>O. tenuiflorum</i>	✓
B13		<i>O. tenuiflorum</i>	✓
B15		<i>O. tenuiflorum</i>	✓
B16		<i>O. basilicum</i>	✓

Key: Germany- Botanical Institute Karlsruhe Institute of Technology, Kew- The Royal Botanic Gardens, CAMAG- CAMAG Laboratory, Switzerland, BTG- Biomolecular Technology Group

3.5 Conclusion

Traditionally, morphological features have been used to identify particular traits of plant families and species. The condition of the samples, plant age and manufacturing process could affect the plant sample and therefore the plant identification process, too (Jacono, 2009). Similarly, all these factors affect the chemical analysis results as well (McCutcheon, 2002). Compared to all these techniques, DNA analysis has its unique advantage as it is not affected by the plant conditions, age, manufacturing process or growth conditions.

Using the *trnH-psbA* barcode region two species, *O. tenuiflorum* (Ram and Shyam which are native to Indian subcontinent) and *O. gratissimum* (Vrinda which is native to Africa, Madagascar, New Guinea) were identified which were discriminated by several base differences. The high-quality sequence data of the *trnH-psbA* barcode region supported these findings. The final outcome of this study has showed the possibility of having morphologically identical *Ocimum* species which have significantly different genetic barcode (Sáez and Lozano, 2005).

After authenticating unknown samples as *Ocimum*, they were further used to identify *OMT* gene region. *OMT* genes play an important part in certain phenylpropanoids pathway steps and their study is a key aspect of this project. Their sequence study is covered in the next chapter.

3.5.1 Further work

3.5.1.1 Storage for further processes

There are various research projects evolving around DNA analysis which require suitable plant material storage techniques and facilities. Different organisations including sample bank, botanical gardens follow well designed protocols to store plant materials. When samples are collected from wild and remote environments, they need to be organised suitably (Harris, 1994). The specimens can be stored in the form of herbarium voucher, for further study and examination. The herbarium voucher ideally contains all plant parts including above ground (stem, leaves, flowers, fruits) along with below ground structures (mainly roots). The sample is usually kept under several layers of paper and porous wooden sheet for air circulation. The voucher specimen sheet should contain all details about plant including plant's full scientific name, collector name, collection location, time. It would be advisable to store these specimens at recognised institute where appropriate facilities are available for long term maintenance.

Fresh plant samples were freeze dried using liquid nitrogen for long term usage. Dry plant material was grounded and converted into powder form. This was stored in small jars and used regularly for DNA extraction.

4 DNA and Amino Acid sequence analysis of *O-methyltransferase* genes

4.1 Introduction

Sequence analysis involves the detailed study of DNA (Deoxyribonucleic acid), RNA (Ribonucleic acid) and Amino acid (AA) sequences of particular genes to understand their structure, functions, features and evolutionary history. When many sequences are added in databases, the key procedure is to compare the new sequences to the existing ones. This helps to understand the mechanism of the target gene's role in relevant organism.

4.1.1 Methylation of biological molecules

The methylation of biological molecules is carried out with the help of highly specific enzyme called methyltransferases. These enzymes transfer a methyl group to the target molecule. *O-methyltransferases* (OMTs) transfer a methyl group to an oxygen atom in the target molecule. In plants, OMTs are a large family of enzymes that can methylate a wide variety of secondary metabolites including flavonoids, phenylpropanoids and alkaloids. The methylation of phenolic compounds by members of the OMT enzyme group is a prominent part of this research study.

Plant phenolic compounds are one of the most ubiquitous groups of secondary metabolites. They are well known for their antioxidant properties and form an integral part of human diet (Hahlbrock and Scheel, 1989). They are known to have multifunctional roles in plants, among which the role of plant-microbe interaction is prominent. These secondary metabolite productions are markedly increased during pathogen attacks, productions that include the upregulation of genes in the phenylpropanoid pathway.

Eugenol and chavicol, naturally occurring plant-based phenolic compounds, are central part of this entire project. Methyl eugenol and estragole are O-methylated compounds that are derived from eugenol and chavicol respectively. The understanding of the phenylpropanoid pathways is still incomplete. The corresponding enzymes, *eugenol O-methyltransferase (EOMT)* and *chavicol O-methyltransferase (CVOMT)*, involved in this pathway have the highest expression levels in the peltate glands of leaves, where they produce volatile oils.

The study revealed that *Ocimum* species can convert eugenol and chavicol phenylpropanoids into their respective methyl derivatives. The *EOMT* and *CVOMT* cDNA sequence alignments of *O. basilicum* sequences are used as a reference for this study. Based on alignment of the cDNA sequences, these two cDNA encoded proteins are found to be 90 per cent identical. A single AA difference was found to be responsible for the difference in substrate discrimination. Based on this finding, the DNA and amino acid sequence analysis of *EOMT* and *CVOMT* in various *Ocimum* species was carried out using several bioinformatics tools.

Some details regarding the synthesis of key phenylpropanoids including eugenol, chavicol, ME and estragole are available (Anand *et al.*, 2016). Two target OMT enzymes for this project, *EOMT* and *CVOMT*, have successfully been identified in sweet basil (*O. basilicum*) crude protein extracts, which were able to produce two methyl derivatives, ME and estragole, from eugenol and chavicol phenylpropanoids respectively. SAM (S-adenosylmethionine) acts as the methyl donor during this reaction (Figure 4-1).

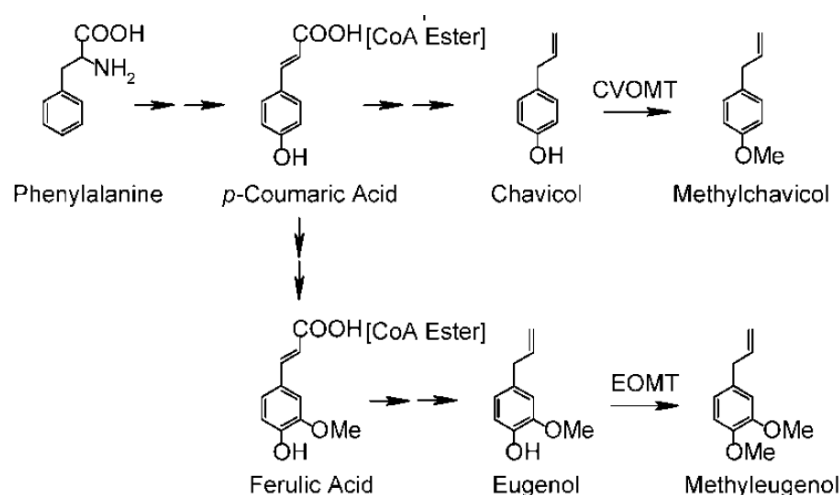


Figure 4-1: Secondary metabolite pathway leading to methyleugenol and methylchavicol in *O. basilicum*. Single arrows indicate the catalytic reaction of *EOMT* and *CVOMT* enzymes, double arrows indicate multiple conversions (Gang *et al.*, 2002)

According to Gang *et al.* (2002), two *OMT* enzymes, *EOMT* and *CVOMT*, were identified in crude protein extracts of sweet basil, though the ratio of these two activities was dissimilar in different chemotypes of *O. basilicum*. Their activity was also discovered to be at its highest level in glandular trichomes present on young leaves and stems. In this way, the methylation of two different phenylpropanoids s two distinct, albeit very similar, enzymes.

4.1.2 Gene expression

The synthesis of any functional gene product is the result of the process called ‘gene expression’. The functional gene often codes for a protein, although it can be a functional RNA such as transfer RNA and small nuclear RNA. Gene expression is an essential process for the synthesis of a functional gene product using information from the gene. This process can be seen in all organisms including viruses, prokaryotes and eukaryotes. Gene regulation can control cell structure function, differentiation and adaptability in an organism. The changes in gene sequences can serve as a substrate to study the evolutionary changes that eventually affect the gene’s function within the cell. Gene activity and its regulation thus profoundly affects the development of individual organisms.

A gene expression process contains several steps that can be divided into the two main categories of transcription and translation.

4.1.2.1 Plant *O*-methyltransferase enzymes

As discussed earlier, a large enzyme family is present in plant organism that methylate several secondary metabolites including flavonoids and phenylpropanoids. This group of enzymes act as a structural support to cell walls. The main aim of this study is the analysis of the target *OMT* proteins that include two groups of enzymes, *EOMT* and *CVOMT*.

4.1.2.1.1 Eugenol *O*-methyltransferase enzyme

Eugenol and their methyl derivatives are ubiquitous in the plant kingdom, especially in the species of the Lamiaceae family. Their pungent smell makes them widely popular in the fragrance and culinary industries. *EOMT* is a common enzyme that uses S-adenosyl-L-methionine ($C_{15}H_{22}N_6O_5S$) as the methyl donor during the methylation process of these secondary metabolites. The size of this enzymatic protein is around 40 kDA, and it is generally stable below 36°C (Wang and Pichersky, 1998). According to Renu *et al.* (2014), the *O. tenuiflorum* *EOMT* genomic DNA study revealed the presence of the open reading frame and intron, 843bp and 90bp long respectively. The enzyme expression level is at its highest in young flowering plants, being reduced in the mature and post-flowering stages. This finding supports the different levels of secondary metabolites in different *Ocimum* chemotypes (Renu *et al.*, 2014).

4.1.2.1.2 Chavicol *O*-methyltransferase enzyme

S-adenosyl-L-methionine (SAM) is the same methyl donor used by the chavicol *O*-methyltransferase enzyme during the methylation process in which secondary chavicol metabolites form estragole (methyl chavicol). There are striking similarities between *CVOMT* and *EOMT* enzymes. The production of *CVOMT* is also abundant in young *O. basilicum* tissues and lower in mature ones. The expression level of *CVOMT* is high in certain genotypes, and so consequently is the estragole level, while genotypes combining *EOMT* and *CVOMT* demonstrate the production of ME and Estragole respectively. These results indicate the presence of two distinct enzymes on *Ocimum*

chemotypes, one specific to ME and another that acts on chavicol during the methylation process (Lewinsohn *et al.*, 2000).

Like other enzymes, an activity of these two *O-methyltransferase* enzymes depends on the availability of substrates and their specificity in various *Ocimum* species and their relevant chemotypes.

4.1.3 Plant Genomics and Bioinformatics

A sequence study is a convenient means of studying plant genes and the characteristics of valuable plant species in detail (Feuillet *et al.*, 2011; Govindaraj *et al.*, 2015). The availability of, and updated information on, plant genomes creates a fruitful environment within the world of plant research. It can also create strong relationships between research groups across the globe (Raes and Bork, 2008). The sequence study marks a new era in the form of Genomics, which began with ‘Sanger sequencing’, and is now gradually being replaced by new ‘High throughput’ technology (Siegel and Saukko, 2012). The availability of genome sequence data played a crucial role in the discovery of novel genes. A complete genome study of *Arabidopsis*, then rice plants, inspired scientists to combine plants species of interest with new sequencing techniques to devise their novel gene projects (Meinke, *et al.*, 1998; Edwards and Batley, 2004). Of the various online databases, CoGepedia has recorded the sequences of around 100 plant species and contains some pending sequence details of projects that are still incomplete (Pagani *et al.*, 2012). This data provides unique and valuable information that reveals interesting facts about critical genes and their pathway details (Michael and VanBuren, 2015). *Arabidopsis* and other plant species, the study of whose genomes has been completed, are now used as model organisms for the study of new genes (Meinke, *et al.*, 1998). The sensible use of available genes sequences can improve plants’ quality and can eventually be used for the well-being of humans.

These technologies have led to a number of new research projects, revolving around gene sequences, that require common locations, called ‘databases’ to store newly invented sequences. Databases are either freely accessible public bodies or private databases that require personal logins. Both are managed by certain organisations that

monitor the data available for use by researchers in their genome projects. NCBI, the DNA Data Bank of Japan and GenBank are well known examples of public databases. Database entries have increased dramatically in number since 2004 (Ong *et al.*, 2016). NCBI's total entry numbers currently stands at 5,000,000 (RefSeq, 2015), with 35,000 entries being added monthly. As well as general plant sequence databases, certain databases are species-specific. Examples include The Arabidopsis Initiative Resource (TAIR) (Swarbreck *et al.*, 2008), MaizeGDB (Lawrence *et al.*, 2004) and The Rice Annotation Project (RAP-DB) (Sakai *et al.*, 2013). Most of these databases relate to other facilities such as BLAST and Uniprot search engines that help researchers unearth more details about similar sequences while entering their experimental dataset into the database.

The sequences deposited in public databases contain very limited information. Ideally, each sequence should contain relevant information (for example, epitopes, functional domains, genetic markers and haplotypes) to obtain complete genetic maps (Michael and VanBuren, 2015). Another major issue regarding the regular updating of gene sequences is that this requires a bridge between database developers and plant researchers (Dhanapal and Govindaraj, 2015). Plant genome databases are thus faced by technical challenges, although their potential to study large collections of plant genes is huge. The exponential increase in plant genome data that will result from the introduction of NGS in sequencing requires suitable tools and data management. Effective analysis methods for genome comparison, phylogenetic investigation to study evolutionary relationships, and most importantly skilled investigators, are indeed required to study plant genomes using bioinformatics tools (Ong *et al.*, 2016). While examining the similar sequences of the *EOMT* region from different species during the literature review, these factors greatly affected the research direction.

4.1.3.1 Resolving conflicts after alignment

Aligning many sequences and obtaining intelligible results is a difficult and time-consuming process. Previously, only the small-number of sequences were aligned, these later were merged with large amounts of sequence data by using bioinformatics tools (Notredame *et al.*, 2000). The whole alignment concept was initiated with the pairwise

alignment process in which similarity and gap penalties were summarised (Needleman and Wunsch, 1970). Most multiple alignment programs incorporate a pairwise alignment step. Segment-based alignment (with some modifications), where the first similarity score has been introduced and arranged sequentially, has also been introduced (Subramanian *et al.*, 2008). Whatever the approach taken to align sequences, the main problem of aligning weakly conserved regions that can easily be missed at the first step of the alignment process, still prevails (Kececioglu *et al.*, 2000).

4.1.3.2 Protein importance

Proteins are important molecules in all living organisms. Proteins or polypeptide chains are made up of 20 essential AAs from nucleotide sequences. A group of three nucleotide base pairs encoded for each AA are grouped in a particular order to form AA chains. This is followed by the construction of the folding mechanism (Anthony *et al.*, 1989; Xie and Schultz, 2005). Proteins are responsible for the majority of events including maintenance, protection, reproduction and all DNA mechanisms including replication and multiplication. (Lesk, 2001; Green, 2005). Protein functions can be categorised on the basis of their nature of their involvement in the process. Examples are the structural proteins responsible for cell shape (Cabeen and Jacobs, 2005), mitochondrial proteins that analyse transmembrane events (Bloom, 2005), and defence proteins that recognise foreign objects entering the body and activating an immune system (Delves and Roitt, 2000). Apart from these major functions, various minor mechanisms occur with the help of protein molecules, which is why the study of proteins is so important to understand the naturel processes. By understanding their mechanisms and folding structures, various unanswered questions regarding the human physiology can be revealed. The study of proteins is also extremely useful in therapy, as it can reveal the root causes of diseases and their treatments. Previously, only the principles of applied biology were used in this context, but now knowledge relating to computation, chemistry, physics and molecular studies combine in the field of bioinformatics (Luscombe *et al.*, 2001). As with DNA databanks, databases such as NCBI, Genebank and PDB that contain details of AA and protein structures obtained from research groups working in this field are also available online (Galperin, 2006).

4.2 Material and methods

4.2.1 Primer design

A range of primer pairs from previous publication were used to identify an *EOMT* gene. (Gang *et al.*, 2002; Renu *et al.*, 2014). One external primer pair was designed using ‘Primer Design’ software (see the Results section for more details). The same details were helpful in identifying the *CVOMT* region of different *Ocimum* species (Table 4-1).

4.2.1.1 An internal primer strategy

The *EOMT* amplicons derived with the help of the *EOMT* primer pair were over 1,000 bases long. Their quality was reasonable, the sequences reading from each end, although they did not overlap sufficiently in the middle. These primers were named according to the genus name and location on the gene. For example, *Ocimum* primers covering internal parts of the gene were identified as ‘*Ocimum_Int_F*’ and ‘*Ocimum_Int_R*’ (forward and reverse primers respectively) (Figure 4-5).

Different numbers were assigned if more than one primer was designed for the same region, examples being Int_F1 and Int_F2 (see the results in Figure 4-9). The internal primer pairs Int_F2+ Int_R2 are later renamed as Int-F and Int-R primer pairs to produce a sequence in both directions in order to obtain the full length of the PCR product.

Four external and two internal primer pairs were designed in order to identify the *EOMT* gene can be seen in Table 4-1.

Table 4-1: PCR programme for different primer pairs

Primer	Primer sequence	PCR programme
<i>EOMT</i> -F <i>EOMT</i> -R (0.5 µM)	IGTCGACAGAGCAACTTCTT GGATAAGCCTCTATGAGAGACC	95°C for 4 min, 35 cycles at 94°C for 30 s, 55°C for 1 min, and 72°C for 2 min, followed by a final extension step at 72°C for 10 min
Renu IS-F (Tm 62°C) Renu IS-R (Tm 61°C) (0.5 µM)	TCCCACTTTCACAAACCCAT ACAACATGTGGGAGGTCAATA	95°C for 4 min, 35 cycles at 94°C for 30 s, 59°C for 1 min, and 72°C for 2 min, followed by a final extension step at 72°C for 10 min
Gang OMT 5' (Tm 53°C) Gang OMT 3' (Tm 47°C)	AATGGCATTGCAAAAAGTA CCTTAAATTTGTAGCTATC	Initial denaturation step at 95°C for 7 min; 30 cycles of 95°C for 1 min, 45°C for 30s, 72°C for 1 min; final extension at 72°C for 7 min
Set 2-F (Tm 62°C) Set 5-R (Tm 62°C)	GCCGAGCATGGGTAGATTT CTCATCGTTCCAGTCGTGTATT	Initial denaturation step at 95°C for 7 min; 30 cycles of 95°C for 1 min, 60°C for 30s, 72°C for 1 min; final extension at 72°C for 7 min
Int_F1 and Int_R1	AGTGAGTACATGTGCCACGA GGCACCATATGAGCGAATGG	Initial denaturation step at 95°C for 7 min; 30 cycles of 95°C for 1 min, 53°C for 30s, 72°C for 1 min; final extension at 72°C for 7 min
Int_F2 and Int_R2	ACGACATGTGGGAGGTCAAT GCCGAGCATGGGTAGATTTT	

Among all these primer pairs, *EOMT* primers (Renu *et al.*, 2014) and second internal primer pairs were used to amplify the *EOMT* gene from different *Ocimum* species. Full-length cDNA sequence of *O. basilicum* (*EOMT* and *CVOMT*) were obtained from NCBI databases. Figure 4-5 shows where each primer sits on the *EOMT* gene.

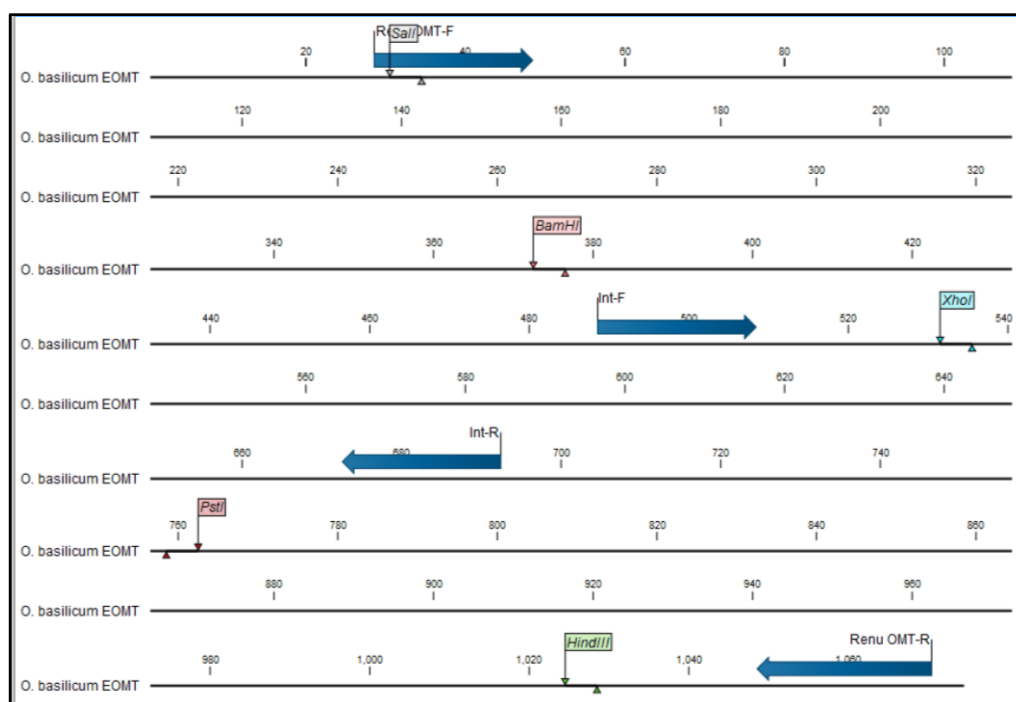


Figure 4-2: Primer design strategy to obtain the details of the full-length *EOMT* gene. The position of *EOMT* and Internal primers can be seen on the *OMT* gene

4.2.2 DNA samples

Many *Ocimum* species are available naturally. Some of their readymade DNA samples were collected by previous research students of the Biomolecular Research Group (BTG) (Bhamra *et al.*, 2016). Their DNA samples have been characterised by using the nuclear ribosomal Internal Transcribed Spacer (ITS) and the chloroplast intergenic spacer *trnH-psbA* regions. The same DNA samples were later used in this study to identify the *OMT* region (Table 4-2).

Table 4-2: BTG DNA samples

Sample No.	Species	Sample No.	Species
A14	<i>O. tenuiflorum</i>	C20	<i>O. tenuiflorum</i>
A17	<i>O. tenuiflorum</i>	C21	<i>O. tenuiflorum</i>
A18	<i>O. tenuiflorum</i>	C22	<i>O. tenuiflorum</i>
A19	<i>O. tenuiflorum</i>	C23	<i>O. tenuiflorum</i>
B01	<i>O. tenuiflorum</i>	C24	<i>O. tenuiflorum</i>
B02	<i>O. tenuiflorum</i>	C25	<i>O. tenuiflorum</i>
B02-R	<i>O. tenuiflorum</i>	C73	<i>O. tenuiflorum</i>
B02-S	<i>O. tenuiflorum</i>	C74	<i>O. tenuiflorum</i>
B03	<i>O. tenuiflorum</i>	C75	<i>O. tenuiflorum</i>
B04	<i>O. tenuiflorum</i>	C76	<i>O. tenuiflorum</i>
B05	<i>O. tenuiflorum</i>	C77	<i>O. tenuiflorum</i>
B06	<i>O. tenuiflorum</i>	C78	<i>O. tenuiflorum</i>
B06G	<i>O. tenuiflorum</i>	C79	<i>O. tenuiflorum</i>
B07G	<i>O. tenuiflorum</i>	C80	<i>O. tenuiflorum</i>
B08	<i>O. tenuiflorum</i>	C81	<i>O. gratissimum</i>
B09-KP	<i>O. tenuiflorum</i>	C82	<i>O. selloi</i>
B09a	<i>O. tenuiflorum</i>	C85	<i>O. americanum</i>
B09b	<i>O. tenuiflorum</i>	I41	<i>O. tenuiflorum</i>
B09c	<i>O. tenuiflorum</i>	I42	<i>O. tenuiflorum</i>
B09d	<i>O. tenuiflorum</i>	I43	<i>O. tenuiflorum</i>
B09e	<i>O. tenuiflorum</i>	I44	<i>O. tenuiflorum</i>
B13-Ebay	<i>O. tenuiflorum</i>	I45	<i>O. tenuiflorum</i>
B15a	<i>O. tenuiflorum</i>	I46	<i>O. tenuiflorum</i>
B15b	<i>O. tenuiflorum</i>	I47	<i>O. tenuiflorum</i>
B16	<i>O. basilicum</i>	I48	<i>O. tenuiflorum</i>
B101	<i>O. tenuiflorum</i>	I49	<i>O. tenuiflorum</i>
B102a	<i>O. tenuiflorum</i>	L10 Met num	<i>O. tenuiflorum</i>
B103	<i>O. tenuiflorum</i>	L11-Met ant	<i>O. tenuiflorum</i>
B104	<i>O. tenuiflorum</i>	Melton Rd	<i>Tulsi Tablet</i>
7810	<i>O. kilimandscharicum</i>	Pantadai	<i>Ocimum spp</i>
8285-58	<i>O. tenuiflorum</i>		
5748	<i>O. citriodorum</i>		

4.2.3 Use of Bioinformatics tools

NCBI BLAST (<https://www.ncbi.nlm.nih.gov/BLAST/>) is a widely-used tool for comparing a variety of biological sequence details including nucleotides of DNA and AA of protein sequences. As several *Ocimum* sequences were obtained during the study, an alignment tool of CLC Main Workbench was used throughout. CLC is the main software used to analyse all experiment data. CLC is a helpful resource for calculating the differences and variability scores within an alignment. Around 30 sequences of genomic DNA followed by AA sequences were studied by using its alignment tools. All manner of reads – forward, reverse and internal – were assembled by selecting the default option to prepare contigs. The same sequences were used to prepare the phylogenetic tree using ‘Phylogeny.fr’ software. Different *Ocimum* species were manually inserted online to obtain the result.

4.2.3.1 Published sequences from databases

A comprehensive list of all the *Ocimum* sequences including *EOMT* and *CVOMT* reference sequences were obtained from online sequence libraries (NCBI and Genebank) (Table 4-3).

Table 4-3: List of accession numbers used throughout the study

Sr No.	Species	Accession No.	Source
1	<i>O. tenuiflorum</i> (Ot1)	EU622042	Renu <i>et al.</i> , 2014
2	<i>O. tenuiflorum</i> (Ot2)	EU622043	
3	<i>O. tenuiflorum</i> (Ot3)	EU622044	
4	<i>O. tenuiflorum</i> (Ot5)	EU622045	
5	<i>O. tenuiflorum</i> (Ot6)	EU622046	
6	<i>O. tenuiflorum</i> (Ot7)	EU622047	
7	<i>O. tenuiflorum</i> (Ot8)	EU622048	
8	<i>O. tenuiflorum</i> (Ot2 cDNA)	EU622049	
9	<i>O. basilicum</i> CVOMT1	AF435007	Gang <i>et al.</i> , 2002
10	<i>O. basilicum</i> EOMT1	AF435008	

4.2.4 CLC software

The sequence data generated by Sanger sequencing were analysed using CLC Main work bench. The sequencing data obtained from the MacroGen website were imported in the CLC 'Navigation area' which can be seen in 'View area'. The data in the form of various files (.abi/.ab1/.clc/.scf/.phd) can be organised into categories by keeping them in separate folders (Figure 4-6).

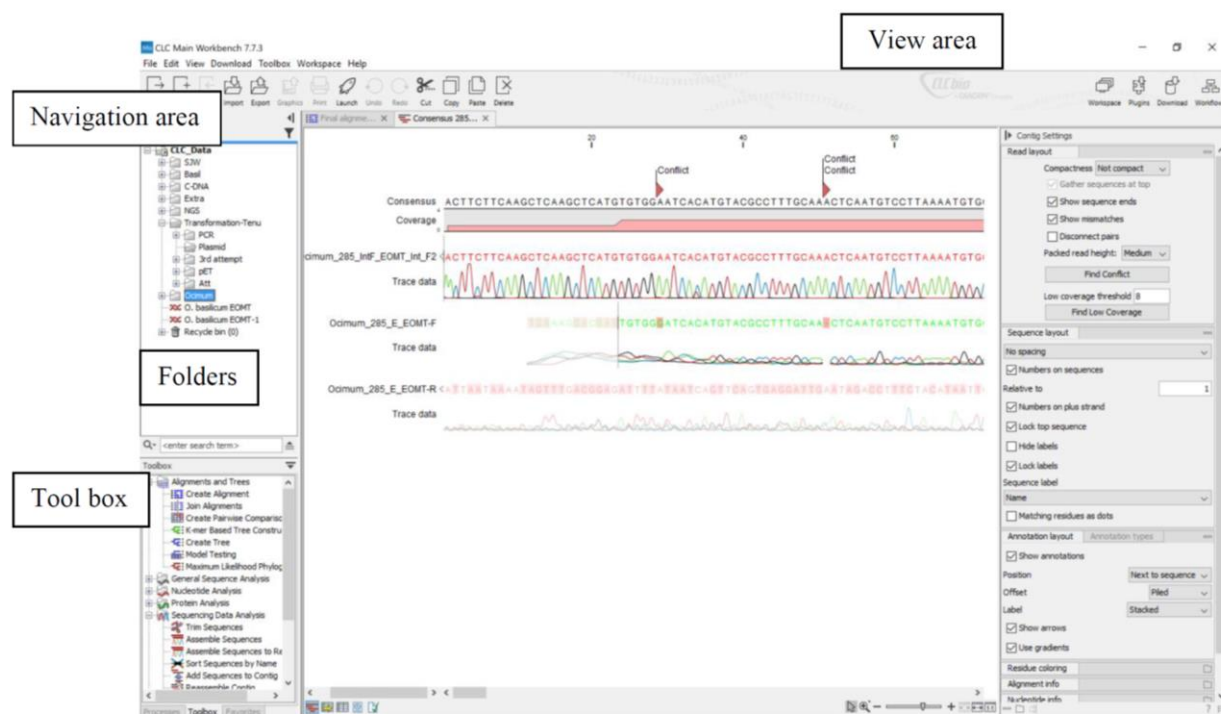


Figure 4-3: The CLC Main workbench showing view, navigation and toolbox areas along with the organisation of data into folders

4.2.4.1 Sequence assembly

CLC Workbench (CLC bio) is the main software used throughout the study. Along with assembling sequences, the study also carried out sequence trimming, alignment and contig preparation using CLC Workbench. Contig involved in consensus sequence preparation using DNA overlapping sequences, were prepared using the default setting with automated vote-based conflict resolution. In case, when contigs could not be prepared using default settings, the alignment stringency was reduced from medium to low (i.e. from 50 bases to 5 bases) level. The reference sequences (AF435007 and AF435008) were also used when needed.

4.2.4.2 Conflict resolution

An *OMT* gene was successfully identified in silico by using *EOMT* primers. At the beginning of the project, the PCR amplification of this region was increased and sequenced to prepare contigs. After assembling relevant reads, the software itself produced the conflicts based on the automated vote settings. These conflicts were resolved one by one by altering the bases or keeping them in their original form (Figure 4-7).

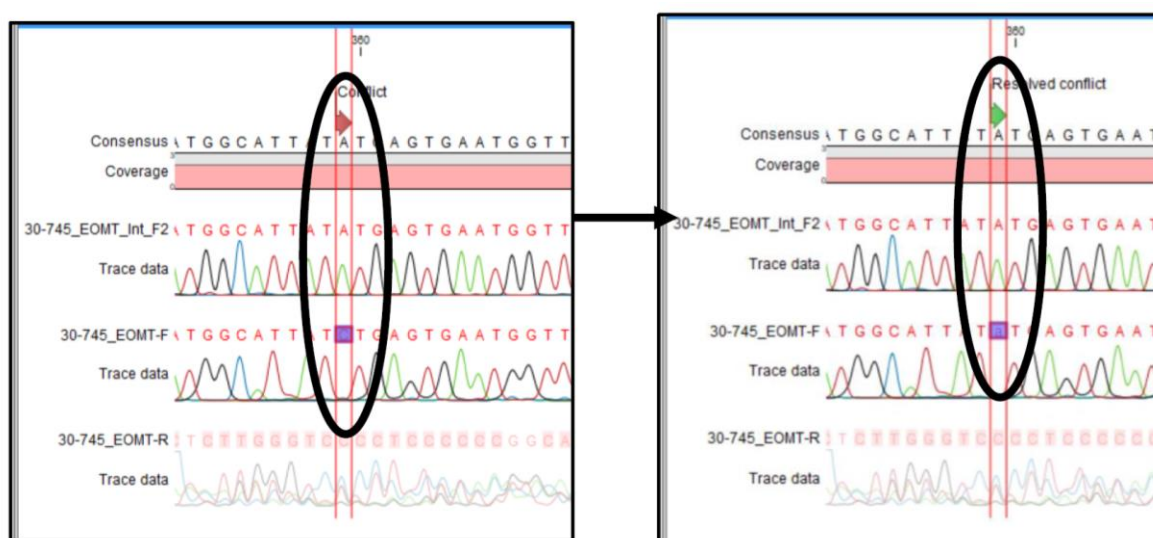


Figure 4-4: Process of resolving conflicts. The conflict annotation in the graphical view. The above example shows a conflict at position number 359 in which the Int-F strand shows a base call of “A” and the EOMT-F strand shows a ‘G’ in that position. The conflict was resolved by changing ‘G’ to ‘a’ in EOMT Forward strand.

4.2.4.3 Nucleotide alignment

For comparison purposes, the query DNA sequences were aligned with reference sequences using the CLC alignment facility. Later, by using CLC Main Workbench, the query sequences were studied in detail with the help of *O. basilicum* reference sequences.

4.2.4.4 The translation of a nucleotide sequence in to an amino acid sequence

After determining the nucleotide sequences of different *Ocimum* species, AA sequences were prepared. Different AA frames were aligned (using CLC alignment and Nucleotide analysis>Translate to protein) with the reference sequence (AF435007 and AF435008)

with experiment data to determine the correct AA reading frame (the longest open reading frame). This method was also used to find an intron region. As published cDNA sequences of *Ocimum* were used as reference points, it has been easier to identify an intron region in gDNA. By comparing cDNA and gDNA sequences, the exon-intron boundaries can be seen clearly. While studying any GOI, the identification of intron-exon boundaries helps to understand how mRNAs are produced from gDNA sequences, a process that plays the vital role in the production of the final protein product.

More details about introns can be found in the Appendix section.

4.3 Result

4.3.1 Primers

A range of primer pairs was designed to identify the O-methyltransferase genes from the *Ocimum* species.

4.3.1.1 External (Published) primers

As discussed in the previous chapter (Ch 3 Authentication), the genomic DNA samples of various *Ocimum* species were authenticated using the *trnH-psbA* primer pair. The positive DNA samples were further used to identify the *EOMT* gene, and four specific primers (as shown in table 4-1) were therefore constructed.

Among all four primer pairs, EOMT-F and EOMT-R (Figure 4-8) returned positive results for the *Ocimum* plant samples that were collected during the first year of the PhD from the local Hindu community in Leicester. After detailed conversations with South Asian people, three Tulsi types, *Ram*, *Shyam* and *Vrinda*, were used for the first round of DNA extractions, being used to check the efficiency of the primer pairs. These samples were authenticated by using the *trnH-psbA* primer pair and were identified as *O. tenuiflorum* (*Ram* and *Shyam*) and *O. gratissimum* (*Vrinda*) respectively (See chapter 3 results for more details).

The *EOMT* primer pair was designed by obtaining the appropriate accession numbers of *EOMT* genes from the GeneBank database and using the BLASTn program to check against available sequences (Renu *et al.*, 2014). Based on these results, the *EOMT* primer pair was selected for further research work to identify OMT genes in *Ocimum*.

Figure 4-8 represents PCR results for (a) EOMT-F & EOMT-R (1500 bp) (b) IS-F & IS-R (300 bp) (c) Gang-F & Gang-R (no bands) and (d) Set-2 & Set-5 (no bands), where the three bands that can be seen in Lanes 1, 2 and 3 of Figure 4-8(a) represent *O. tenuiflorum* (Ram), *O. tenuiflorum* (Shyam) and *O. gratissimum* (Vrinda) respectively. Figure 4-8(b) represents faint bands Lane 2 and 3 for *O. tenuiflorum* (Shyam) and *O. gratissimum* (Vrinda) respectively. No bands were present in Figures 4-8(c) and (d). The expected product sizes were only presented by the EOMT-F and EOMT-R primer pairs in Figure 4-8(a).

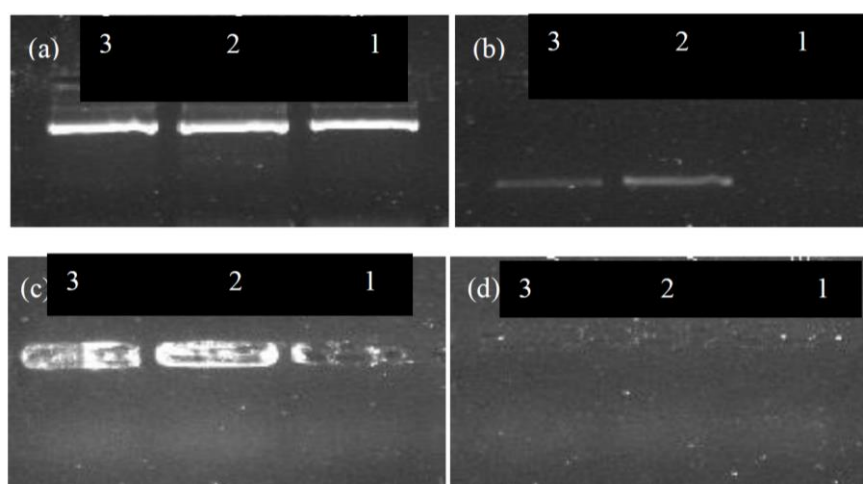


Figure 4-5: Specific primer pair PCR attempt with three *Ocimum* varieties: Samples 1 (Ram) and 2 (Shyam) (*O. tenuiflorum*), and Sample 3 (Wild) (*O. gratissimum*) (A) EOMT-F and EOMT-R (1500 bp) (B) IS-F & IS-R (300 bp) (C) Gang-F & Gang-R (no bands) and (D) Set-2 & Set-5 (no bands). Lane 1: *O. tenuiflorum* (Ram); Lane 2: *O. tenuiflorum* (Shyam); Lane 3: *O. gratissimum* (Vrinda)

The new DNA samples were tested the same way with the *EOMT* primer pair. As shown in the Results section, most of them returned positive results.

The *EOMT* amplicons derived with the help of this *EOMT* primer pair were over 1,000 bases long. They showed reasonable quality sequence reads from each end that overlapped well in the middle but were not long enough to obtain a complete consensus sequence. Two internal primer pairs were therefore designed. In Figure 4-9, bands were

visible in Lanes 1, 2, 3 and 4 [positive control, *O. tenuiflorum* (Ram) sample, *O. tenuiflorum* (Shyam) and *O. gratissimum* (Wild) respectively]. A faint primer dimer band was visible in Lane 5 (negative control) in Figure 4-9(a), while no bands were visible in Lane 5 (negative control) in Figure 4-9(b). The expected product sizes were present in both gel images. Among these two primer pairs, Int_F2 and Int_R2 were selected over Int_F1 and Int_R1 due to better-quality results. Internal primer pairs Int_F2 and Int_R2 were later renamed as Int-F and Int-R to obtain a sequence in both directions along with the full length of the PCR product.

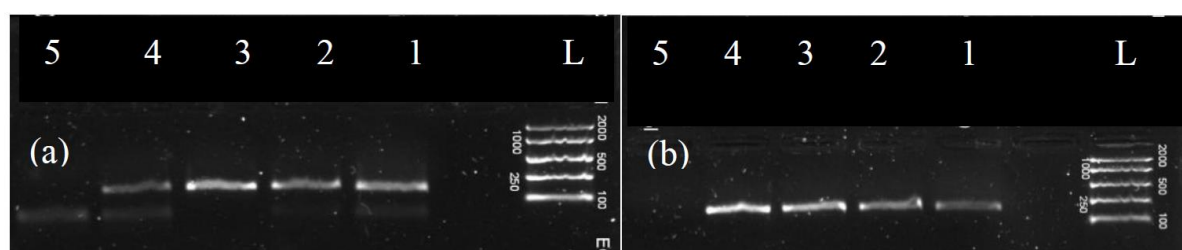


Figure 4-6: Gel results for two sets of internal primers. (A) PCR with Int-F1 and Int_R1. L, DNA ladder. Lane 1: positive control; Lane 2: *O. tenuiflorum* (Ram) sample; Lane 3: *O. tenuiflorum* (Shyam) sample; Lane 4: *O. gratissimum* (Wild) sample; Lane 5: negative control. (b) PCR with Int-F2 and Int_R2. L, DNA ladder. Lane 1: positive control; Lane 2: *O. tenuiflorum* (Ram) sample; Lane 3: *O. tenuiflorum* (Shyam) sample; Lane 4: *O. gratissimum* (Wild) sample; Lane 5: negative control.

4.3.2 Nucleotide alignment

4.3.2.1 Sequence assembly

Trimming, removing conflicts and preparing the consensus sequence were the main steps during sequence assembly. They could be managed using the tool box in the left-hand panel.

Trimming the sequence was the first step in sequence assembly. The poor-quality reads were eliminated through the trimming process. The trimming parameters could be changed, thereby allowing the trimming length and quality to be decided. After trimming, sequence reads were saved with and without the trimmed reads area. The trim annotations were saved separately for future use.

Afterwards, the forward (*EOMT*-F and Int-F) reads were assembled with their relevant reverse (*EOMT*-R and Int-R) reads to prepare the contigs. At this stage the *De novo*

assembly tool was used to align two reads where they overlapped with each other. The alignment option was increased or decreased according to the stringency of the alignment. All these reads including *EOMT*-F, *EOMT*-R, Int-F and Int-R were later assembled together using reference sequences. The sequence assembly parameters were changed from high and medium to low. A higher stringency level avoided more sequence reads and produced short contigs. For the majority of sequence assemblies, lower (an aligned read length of score 5) and medium (an aligned read length of score 50) stringency levels were therefore set, as shown in Figure 4-10.

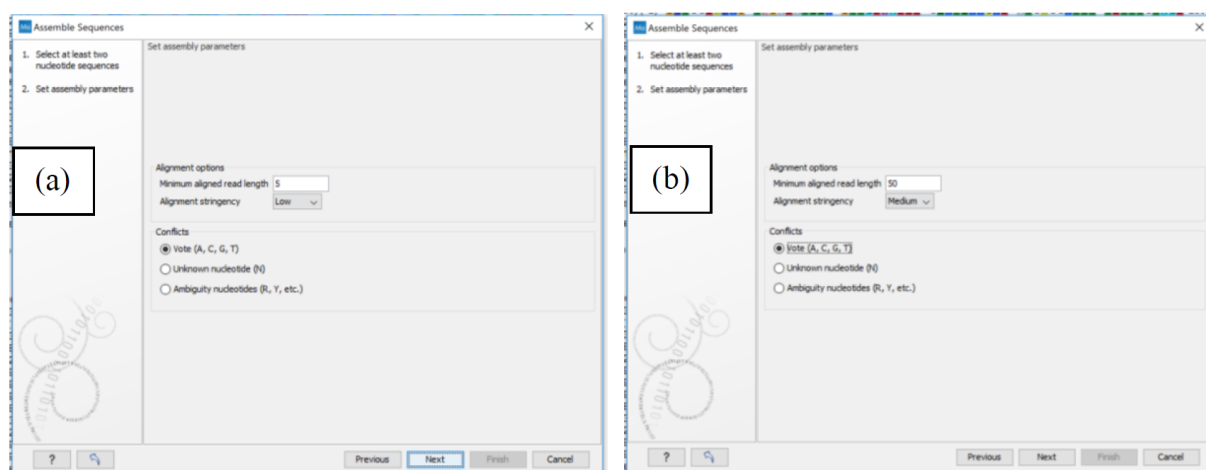


Figure 4-7: Sequence assembly parameter dialogue box. (a) Low stringency level with minimum alignment length score of 5. (b) Medium stringency level with minimum alignment length score of 50.

O. basilicum (*EOMT*) and *O. basilicum* (*CVOMT*) sequences obtained from the NCBI database were used as reference sequences. An example of this assembly result output where conflicts were present can be seen in Figure 4-11.

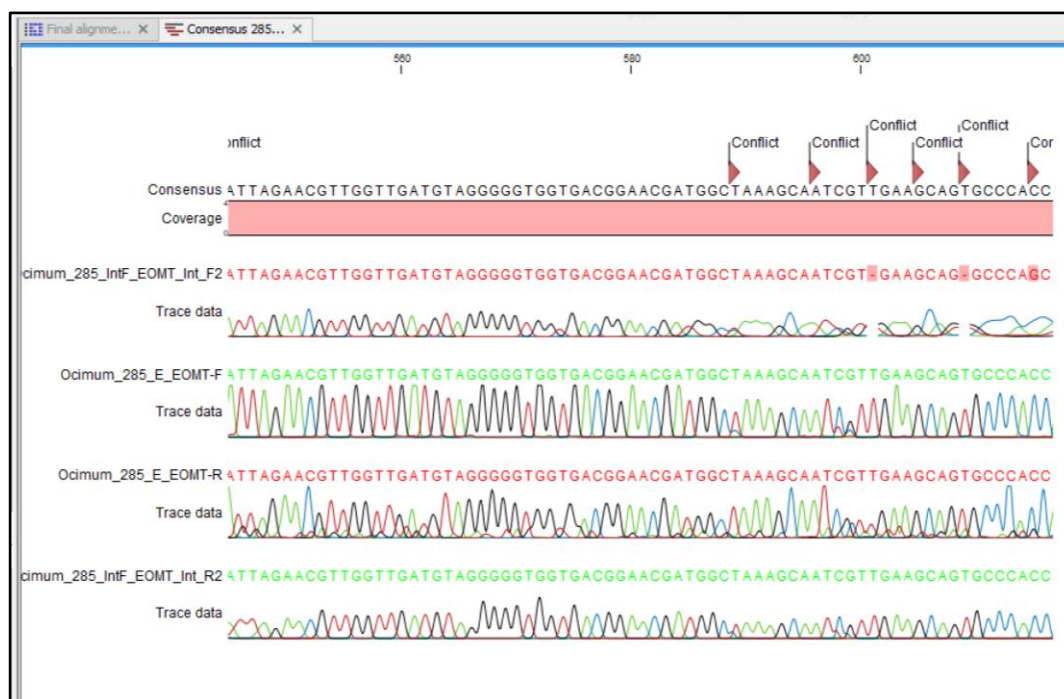


Figure 4-8: Assembly of forward and reverse reads (including *EOMT-F*, *EOMT-R*, *Int-F* and *Int-R*) showing contig preparation with the presence of conflicts.

Conflicts arise where multiple reads overlap and read has mismatched bases at the same position. As shown in Figure 4-12, one conflict has been expanded to improve clarity. In this example, *Int-F* read shows a gap while other reads show the base call of “T”. Before resolving the conflict, the quality of all reads at this position were checked, the *Int-F* quality read being found lower while the other reads including *EOMT-F*, *EOMT-R* and *Int-R* sequence quality peaks could be trusted. Hence the conflict was resolved by replacing the gap with the “T” base in *Int-F* sequence. In this way, each conflict was resolved by voting the correct base (A, C, G, T). To avoid confusion, original contigs and contigs with resolved conflicts were kept in separate folders.

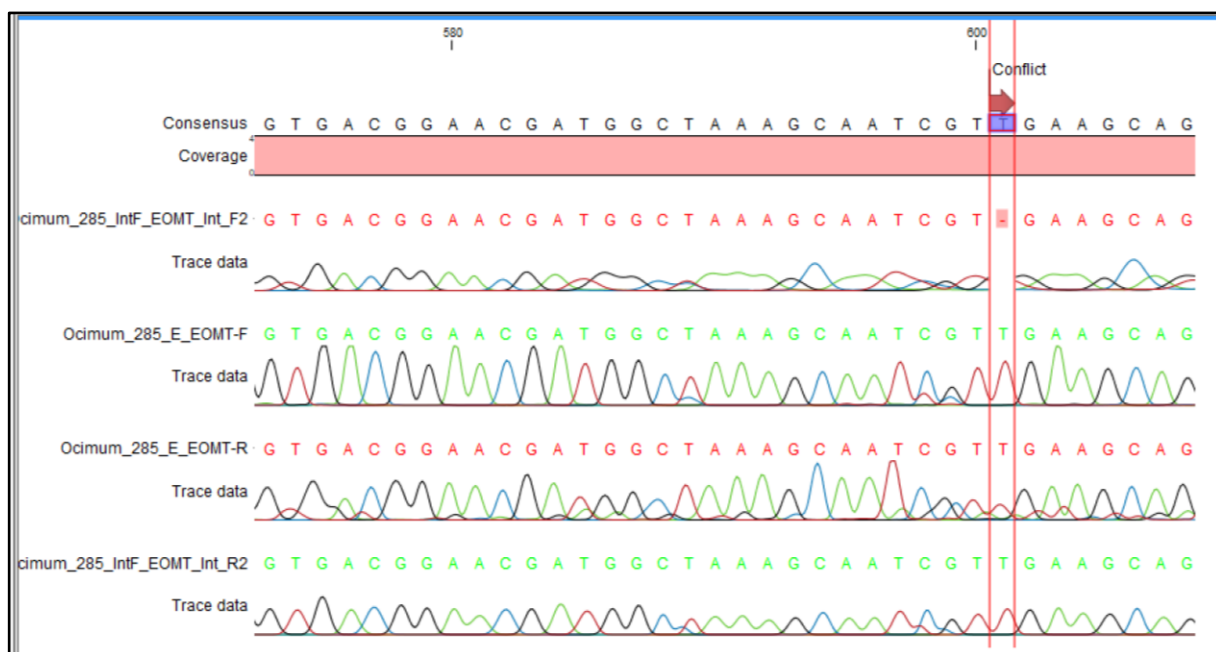


Figure 4-9: Resolving conflicts. The bases' differences must be checked against the reference sequence. From the sequence assembly results of sample 284 (*O. tenuiflorum*), three peaks can be seen at base position 601, where the base was absent on the internal-F sequence. It can therefore be concluded that there should be a T instead of a gap at this base position

When contigs were prepared after resolving all conflicts, the consensus sequence was extracted, saved in a separate folder and used for further analysis.

4.3.2.2 Aligning consensus sequences

Multiple consensus sequences were aligned by using the 'Alignment and Trees' tool. The alignment parameters were changed to increase or decrease the gap settings for making the alignment more stringent. The 'Fast (less accurate)' alignment option was preferable to the 'Short (very accurate)' option, as the former (fast) alignment contains several complicated and very long sequences. An initial alignment contained some conserved regions as shown in Figure 4-13. Gradually, all these regions were carefully studied to prepare the final output of alignment results.

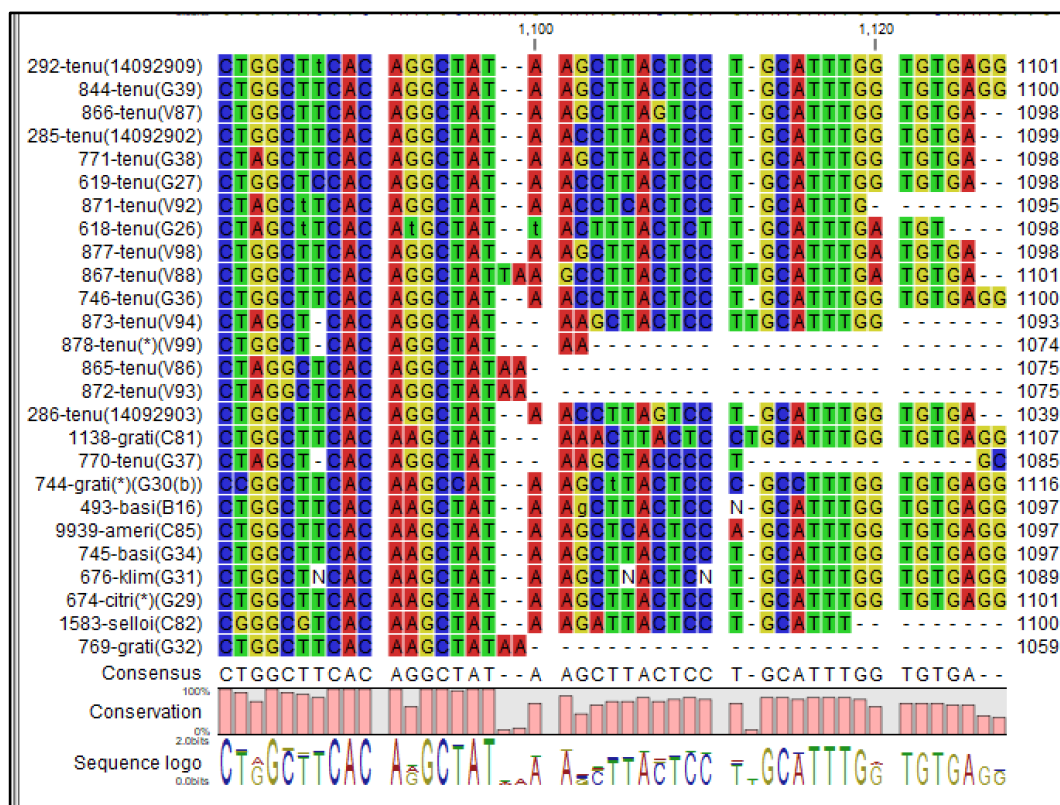


Figure 4-10: An initial alignment containing various *Ocimum* sequences. The alignment was prepared using the CLC alignment tool

The insertion and deletion of a single base in the coding sequence were immediately suspect because they would disrupt the reading frame. The original contig was re-examined in this position to determine whether the sequence reads were consistent with the consensus sequence developing from the alignment (Figure 4-14).

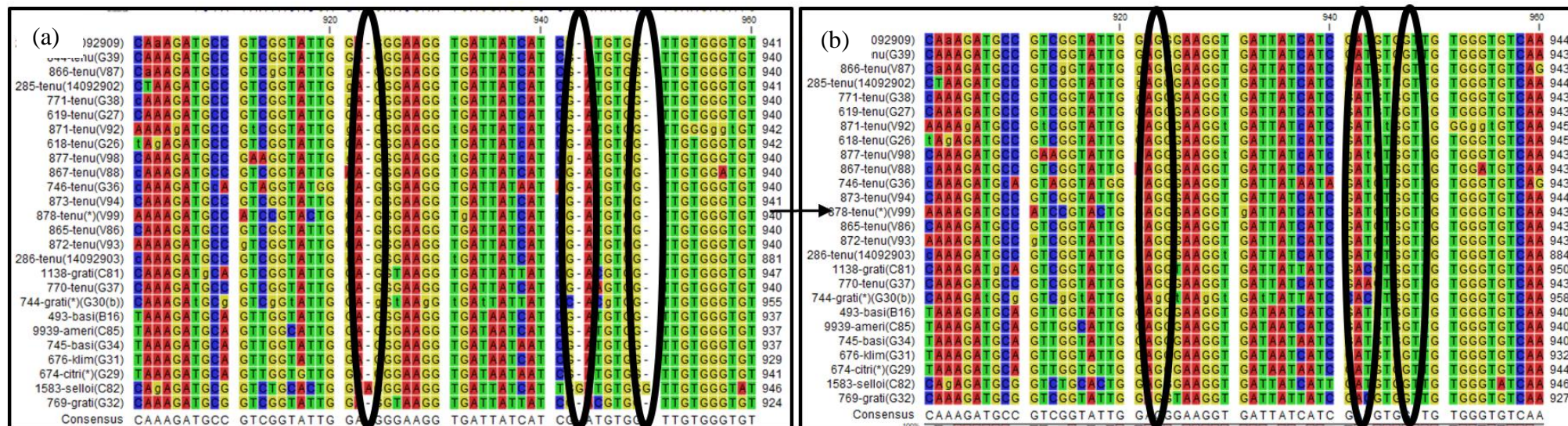


Figure 4-11: A subset of aligned sequences was realigned after removing the gaps. (a) An alignment output represented the gaps presence at nucleotide position no. 923, 943 and 950. (b) These gaps were successfully removed after studying each of the sequences included in an alignment, especially *O. selloi* sequence as it showed an extra base pair which was due to poor sequence quality. The highlighted parts around nucleotide position Nos 922, 942 and 950 did not reveal the presence of gaps, which produced good quality alignment results

As discussed in section 3.2.1, several DNA samples were collected from various sources, of which 26 sequences were successfully sequenced with the full set of *EOMT* primers including *EOMT*-F and *EOMT*-R, Int-F and Int-R. These sequences were trimmed, modified and aligned, as explained above, to prepare contigs. The partial nucleotide alignment sequence can be seen in Figure 4-15. The full alignment can be found in the Appendix.

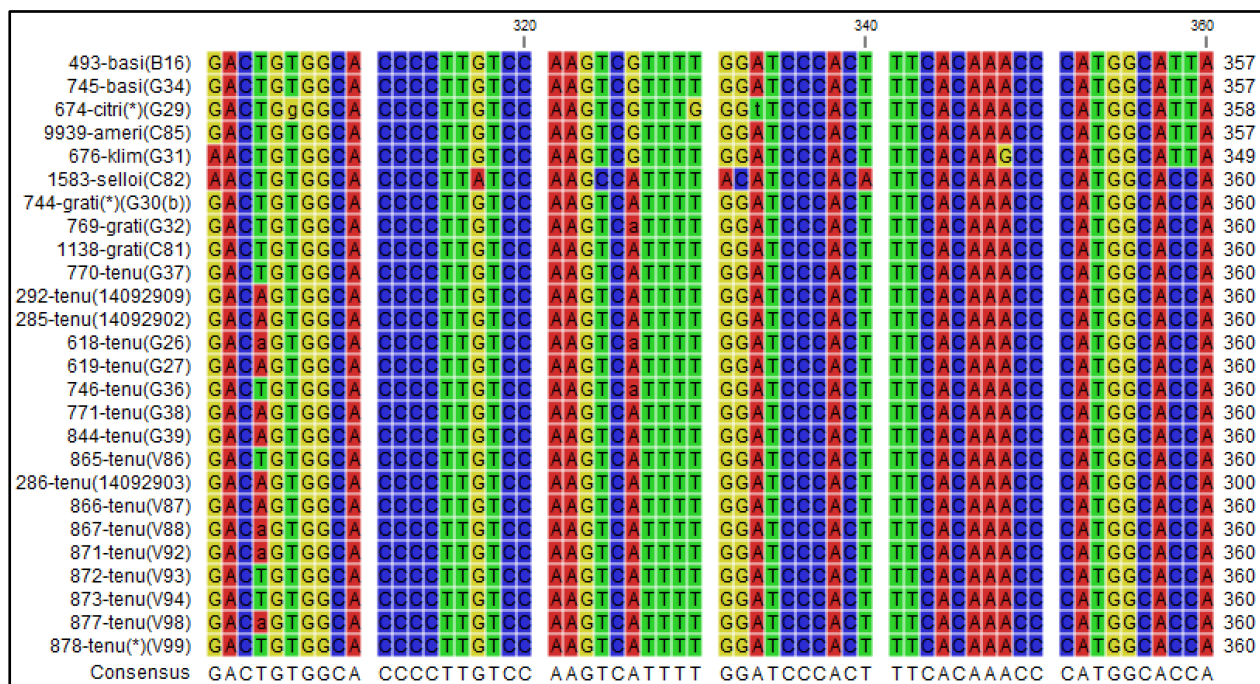


Figure 4-12: A final alignment output. The figures show the partial nucleotide alignment. A complete set of nucleotide alignments is included in Appendix.

4.3.3 Visualising Introns

The best way to identify introns is to align the gDNA sequences with a reference cDNA sequence. The two reference sequences of *O. basilicum* (AF435007 and AF435008) were cDNA sequences. Therefore, apart from resolving conflicts, these reference sequences were also used to identify intron boundaries. Both gDNA and cDNA sequences were carefully studied to identify intron-exon boundaries. The previous research revealed the presence of a 90bp-long intron in *O. tenuiflorum* only (Renu *et al.*, 2014). This project, however, includes several *Ocimum* species and their chemotypes, which required detailed study of all contigs and consensus sequences. As shown in the image, each sequence shows more variation within the intron region, which resulted in different sizes of intron after they were aligned with cDNA reference sequences (Figure 4-16).

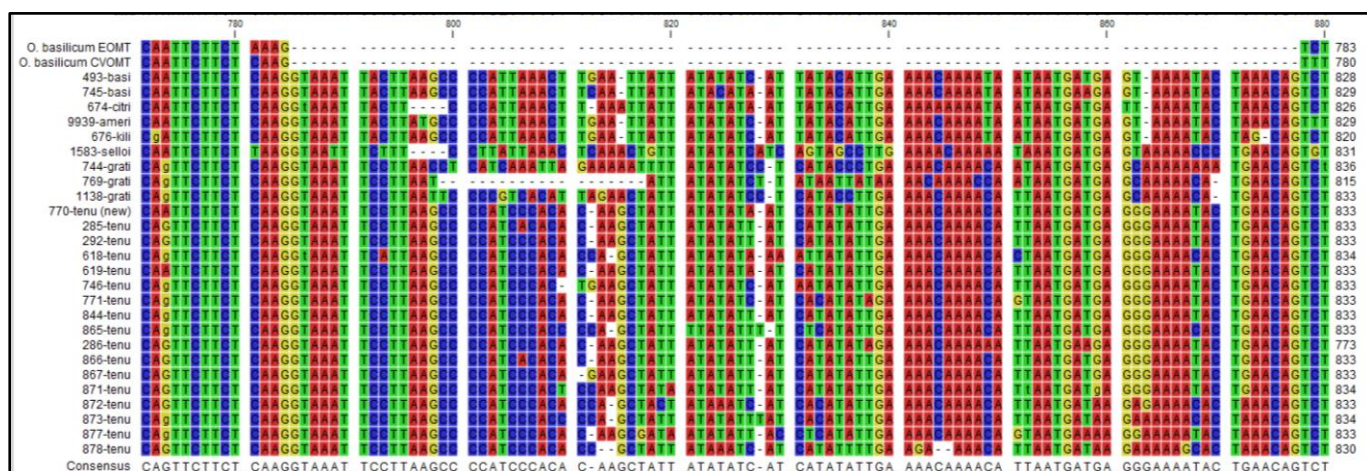


Figure 4-13: Identifying and removing introns from genomic DNA sequences. Introns were missing from two *O. basilicum* (EOMT and CVOMT) reference sequences, which helped identify introns from an experimental set of sequences. Intron-exon boundaries were represented by the bases on either side of each splice site, AG/GT and AG/TC

The intron's presence in various *Ocimum* species was from position 785 to 887 on all gDNA sequences, which accords with *O. basilicum* cDNA reference sequences as shown in MSA in Figure 4-16. The first part of the sequences contains the 'GTAAATTACT' base pair pattern and the final part has the 'TACTAAACAG' bp pattern (with a few exceptions). An entire region of this splice site consensus sequence can be seen in Figure 4-17.

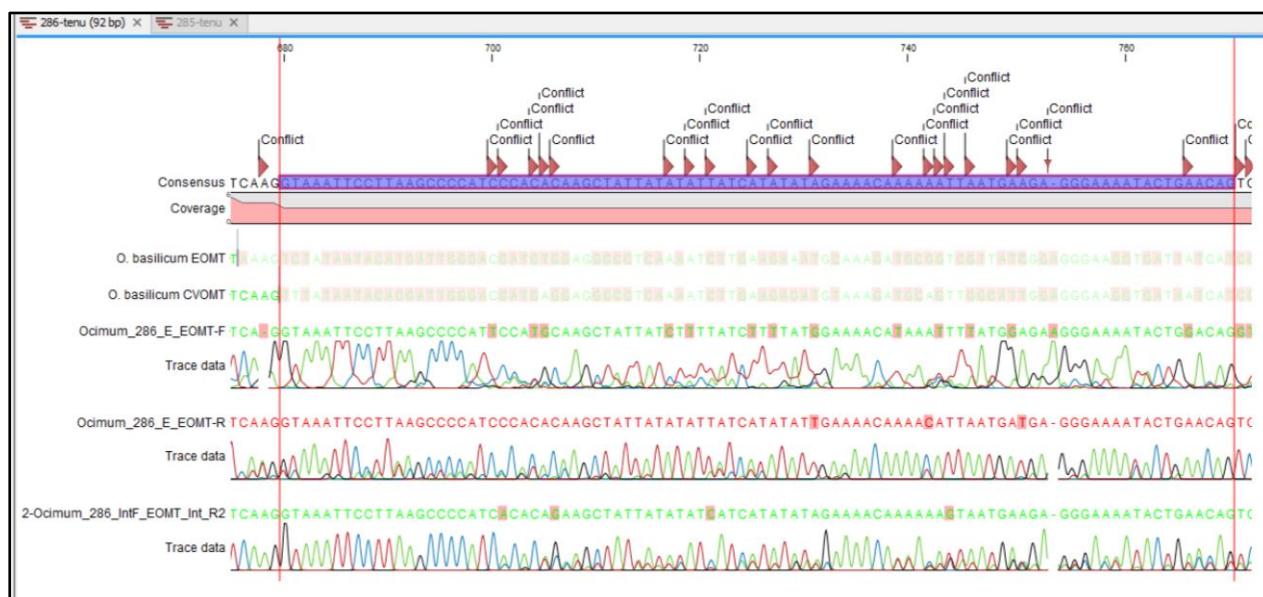


Figure 4-14: An example of the splice site consensus sequence showing intron boundaries from 679 bp to 770 bp.

Each sequence was studied individually to identify intron boundaries and is represented by the ‘AG/GG’ [Figure 4-18 (a)] and ‘AG/TC’ [Figure 4-18 (b)] base pairs. Each consensus sequence was studied individually to identify these boundaries. These differences are highlighted in Figure 4-18.

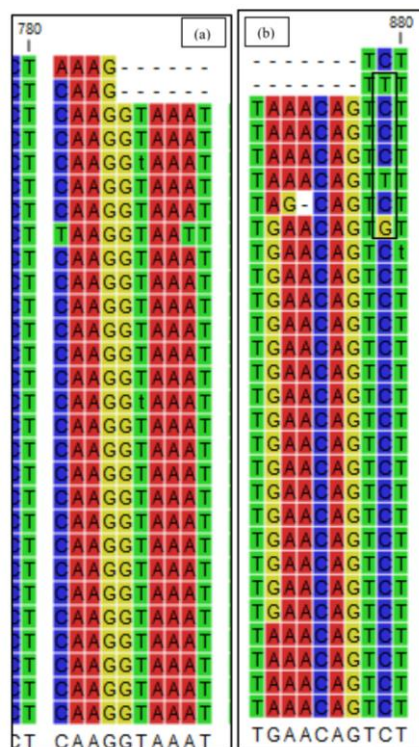


Figure 4-15: Representing intron boundaries by aligning gDNA sequences with cDNA sequences. (a) An exon-intron boundary represented by ‘AG/GT’ base pairs. (b) An intron-exon boundary represented by ‘AG/TC’ base pairs, with a few exceptions highlighted in *O. basilicum* (CVOMT), *O. americanum* and *O. selloi*

As a result, from 74 to 103bp-long introns were found in various *Ocimum* species. Their details can be found in Table 4-4. The most diverse range of intron lengths were found within *O. gratissimum* samples (highlighted in Table 4-4). One explanation of the variation in the length of intron sequences could be that all samples were collected from multiple locations around the world, their details being shown in Table 4-4. The sequences of each species with intron details can be found in the Appendix.

Table 4-4: The details of various bp-length intron present in various *Ocimum* sequences

Sr. No.	<i>Ocimum</i> species and chemovars	Intron bp length	Sample collection place
1	<i>O. basilicum</i> (B16)	91	Akamba garden, UK
2	<i>O. basilicum</i> (G34)	91	Germany
3	<i>O. citriodorum</i> (G29)	93	Germany
4	<i>O. kilimandscharicum</i> (G31)	90	Germany
5	<i>O. selloi</i> (C82)	94	KEW DNA Bank
6	<i>O. americanum</i> (C85)	91	KEW DNA Bank
7	<i>O. gratissimum</i> (G30b)	103	Germany
8	<i>O. gratissimum</i> (C81)	95	KEW DNA Bank
9	<i>O. gratissimum</i> (G32)	74	Germany
10	<i>O. tenuiflorum</i> (14092902)	91	UK
11	<i>O. tenuiflorum</i> (14092903)	91	UK
12	<i>O. tenuiflorum</i> (14092909)	91	UK
13	<i>O. tenuiflorum</i> (G26)	91	Germany
14	<i>O. tenuiflorum</i> (G27)	91	Germany
15	<i>O. tenuiflorum</i> (G36)	91	Germany
16	<i>O. tenuiflorum</i> (G37)	91	Germany
17	<i>O. tenuiflorum</i> (G38)	91	Germany
18	<i>O. tenuiflorum</i> (G39)	91	Germany
19	<i>O. tenuiflorum</i> (V86)	91	CAMAG
20	<i>O. tenuiflorum</i> (V87)	91	CAMAG
21	<i>O. tenuiflorum</i> (V88)	91	CAMAG
22	<i>O. tenuiflorum</i> (V92)	93	CAMAG
23	<i>O. tenuiflorum</i> (V93)	93	CAMAG
24	<i>O. tenuiflorum</i> (V94)	92	CAMAG
25	<i>O. tenuiflorum</i> (V98)	91	CAMAG
26	<i>O. tenuiflorum</i> (V99)	89	CAMAG

Three *O. gratissimum* samples included in study were aligned with *O. basilicum* reference sequences, which clearly showed the discrepancies within their intron sequences (Figure 4-19).

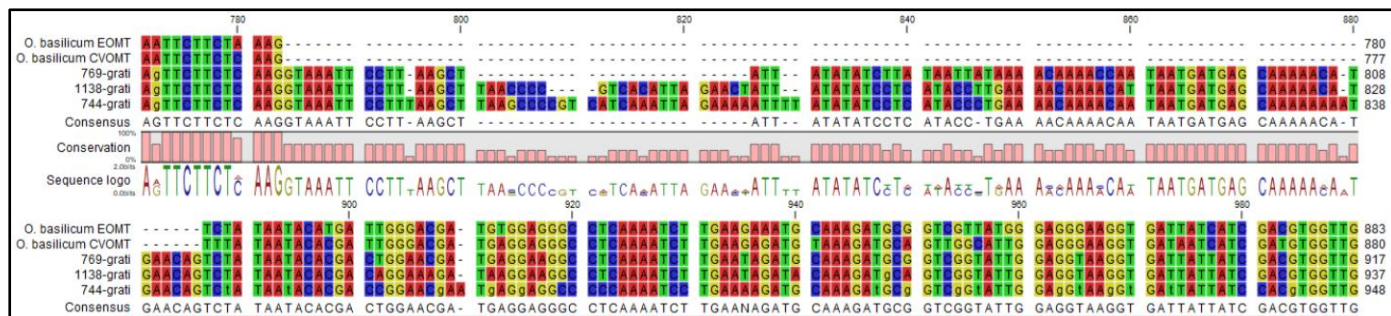


Figure 4-16: Three *O. gratissimum* samples (769, 1138 and 744) were aligned with *O. basilicum* reference sequences shown the most diverse range of intron sequences

4.3.4 AA alignment using nucleotide sequences

Naturally occurring DNA sequences were converted into AA sequences where mRNA plays an intermediary role. These AA sequences determine the formation of related proteins. As

DNA is interpreted in groups of codons (three nucleotides), each DNA sequence has three possible forward (as well as reverse) reading frames, as shown in Figure 4-20.

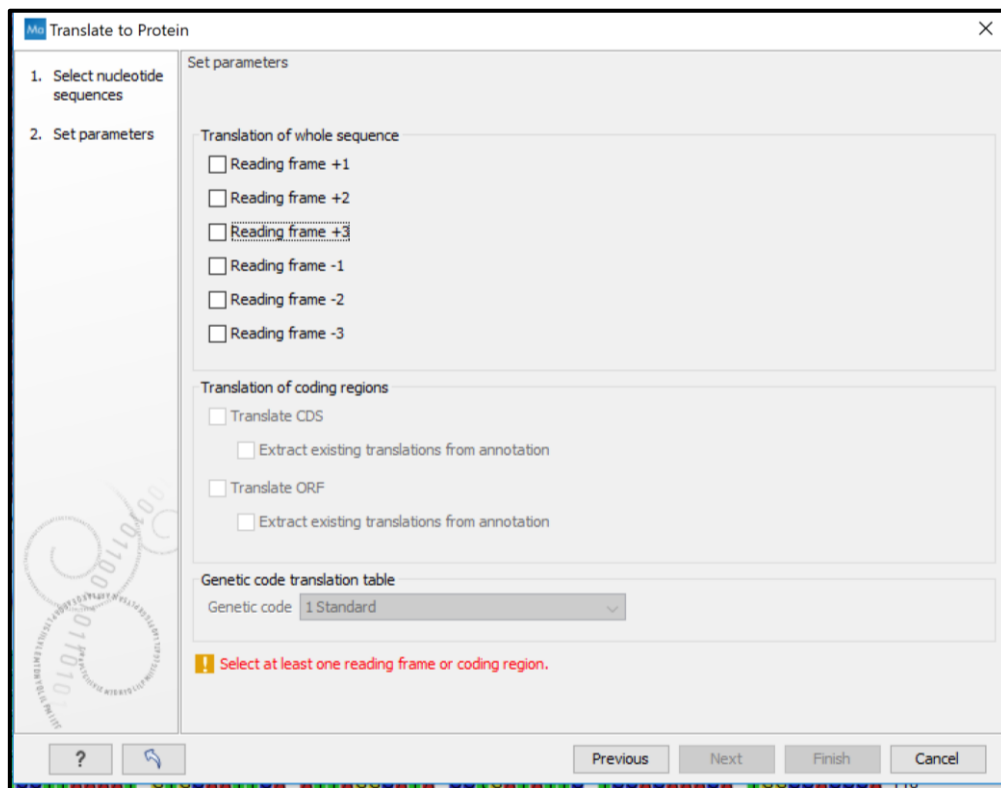


Figure 4-17: Standard translation table showing the possibilities of coding sequence translation frame choices

The +1-reading frame starts from the first nucleotide, the +2 from the second and the +3 from the third. The stop codon is represented by an asterisk insertion at the corresponding position. After nucleotide base alignment, each of them was converted into three forward frames of AA sequences, among which the second frame was selected for AA alignment (Figure 4-21 and Figure 4-22). All this experimental AA sequences were aligned with published reference AA sequences of *O. basilicum* (EOMT) and *O. basilicum* (CVOMT).

		20		40		60	
866-tenu(V87) (+1)	TSSSSSSSCVESHVRLCKLNVLMKNSIRHT*YSPQTWPPNDTFPI						
866-tenu(V87) (+2)	IIQAAHVWNHMYAFANSMISKCAIQIGIPDIIHKHGHPTLSQIIQSIP						
866-tenu(V87) (+3)	FFKLKLMCGITCTPLQTQCP*NVQFN*AYLIFSTNMTQ*HFPNYSNPSPSTKQNP						

Figure 4-18: An example of three forward reading frames from a DNA sequence, among which the second forward reading frame of each nucleotide sample included in the alignment was selected for AA alignment

This AA alignment was edited similarly to the nucleotide alignment, where gaps and AA residues were cross-checked, modified and moved according to the reference as well as other sequences included in the alignment.

(a)

```

618 ACTTCTTCAAGCTCAAGCTCATGTGTGAATCATATGTACGCCTTTGCAAACCTCAATGTCCCTTAAATGTGCAATTCAATTAGGC
100 120 140 160
618 ATACCTGATATTCTCCACAAACATGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCCCCATCAACAAAGCAAATCCC
180 200 220 240
618 AATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAACCTTCTTCATCGAAGAAAAGAACTCTAGTAATCAAGAGGTGCATTA
260 280 300 320 340
618 CTGGCTCACCCAGCCTCGCGCCTCCTCTTGAAGGGGCGCCTTTGACAGTGGCACCCTTGCCAAGTCATTTGATCCCACT
360 380 400 420
618 TTCACAAACCATGGCACCATATGAGCGAATGGTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGCAATGGATGCATGT
440 460 480 500
618 TTTGGGAGAAGTTAGCAAATGAGCGAGCATGGGTAGATTTTTTGTGAAAGTATGAGTTGTGATGCGAGGCTCGTGGCACATGT
520 540 560 580
618 ACTCACTAAGGACTACAAGCATGTGATTGAGGGAATCAGAAGTTGGTTGATGTAGGGGGTGGTACGGAACGATGGCTAAAGCA
600 620 640 660
618 ATCGTTGAAGCAGTGCCACCATGAAATGCACAGTTATTGACCTCCACATGTCGTGGCTGGCTTGGAAAGCACCATAGCTTGA
680 700 720 740 760
618 GCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGCAGTTCTTCTCAAGTCTATAATACGACTGGAACGATGA
780 800 820 840
618 GGAGGGCCTCAAAATCTTGAAGAGATGTAGAGATGCCGTCGGTATTGGAGGAAGGTGATTATCATCGATGTGTTGTGGGTGC
860 880 900 920
618 AACCAGGACGTTGATGAGTTCTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAAGAAAGATTG
940 960 980 1,000
618 AGAAAGAATGGCAAGCTCATCTGATGCTAGCTTCACATGCTATTACTTTACTCTTGCAATTTGATGT

```

Restriction sites: **BamHI** (at position 420), **XbaI** (at position 860), **BglII** (at position 940).

(b)

```

618-tenu(G26) (+2) LLQAQAHVWNHMYAFANSMSLKCAIQLGIPDILHKHGHMPTLSQLLQSIPI NKAKSQCLHRLMRVLVNSNFFIEEK
80 100 120 140
618-tenu(G26) (+2) NSSNQEVHYWLT PASRLLKGA PLTVAPLVQV ILDP TFTP WWHMSEWF THEHHATQFE AANGCMFWEKLANEPSM
160 180 200 220
618-tenu(G26) (+2) GRFFDEAMSCDARLV AHVLT KDYKHV IEGIRTLVDVGGGDGTMAKAI VEAVPTMKCTVIDLPHVVAGLESTDSL SY
240 260 280 300
618-tenu(G26) (+2) IGGDMFRS IPSADAVLLKS I IHDWNDEGLK I LKRCRDAVG I GGVKVI I IDVVGVNQDVDEVLEDQLHFDMMAMY
320
618-tenu(G26) (+2) FNAKERIEKEWAKLRSDASFTCYFTLAFD

```

Figure 4-19: A process to translate nucleotide sequence into AA sequence (A) the nucleotide sequence that codes for (B) the AA sequence (second reading frame)

4.3.4.1 Amino acid sequences alignment

According to Gang *et al.* (2002), *EOMT* and *CVOMT* have evolved from a common ancestral gene to act on different substrates including eugenol and chavicol to produce their methyl derivatives. Through molecular modelling study, one AA difference at position 261 was found to be responsible for substrate discrimination of *EOMT* and *CVOMT* enzymes. This was further confirmed by site-directed mutagenesis.

For a detailed study of AA sequences, each nucleotide sequence was converted to different AA frames. All ORF of gDNA encoded a protein of around 300-400 AA residues. By aligning different AA frames, the correct one was selected for alignment. There were several modifications made after creating the first alignment. The shift in the reading frame occurred because of the incorrect addition of the nucleotide and gaps in the alignment. This type of error can be rectified by inspecting the related codons in properly aligned sequences of nucleotides to determine the location of mistakes, as shown in Figures 4-23 and 4-24.

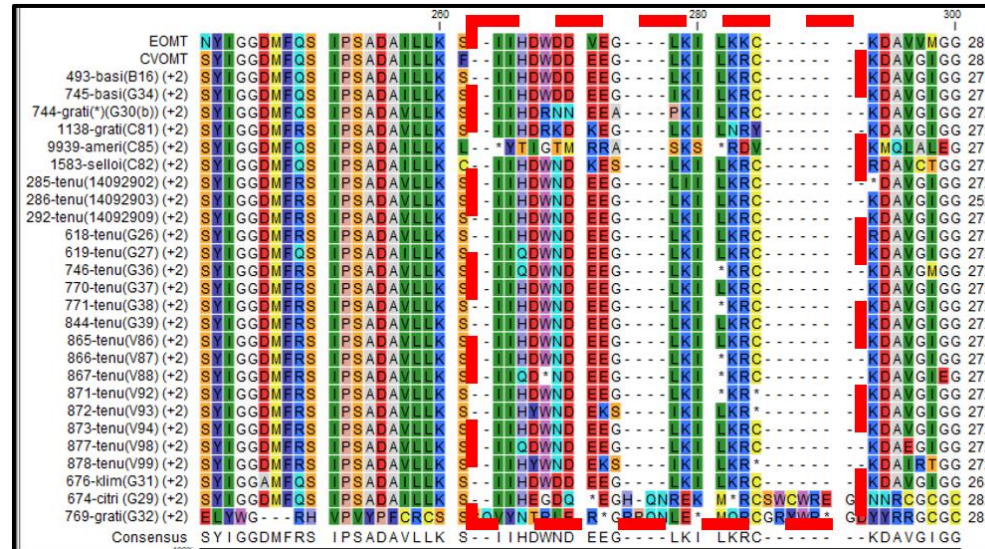


Figure 4-20: Unmodified AA alignment contains frameshift, stop codon and gaps within AA positions no. 262 to 300

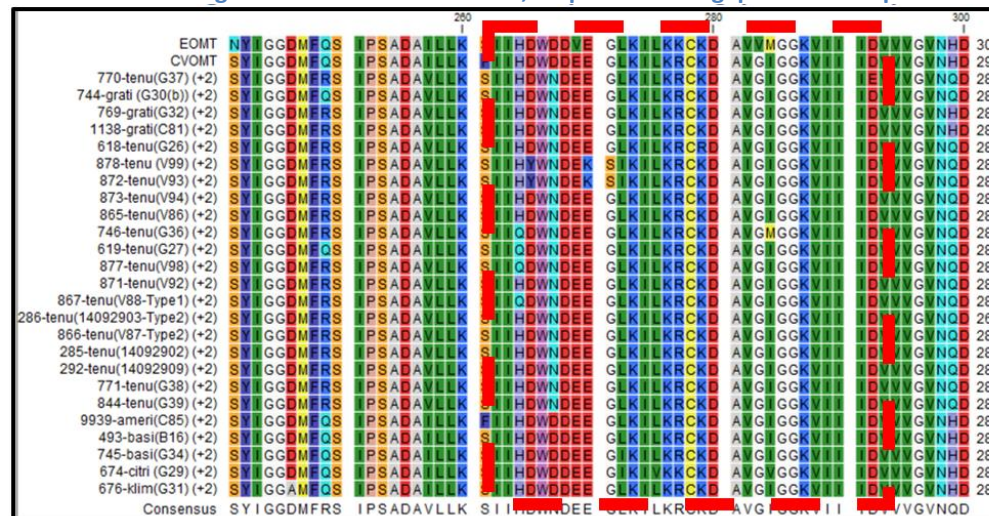


Figure 4-21: Modified AA alignment after removing gaps and cross-checking related nucleotide codons in each coding sequence. The gaps and stop codons from AA positions 262 to 300 were successfully removed to make meaningful alignment

As soon as the alteration was made, the reading frame began to make sense regarding similar or different AA sequences within the variety of *Ocimum* species. While too many sequences were aligned, the results cannot be clearer. Published *O. basilicum* AA sequences were used as a reference during the process of obtaining a clear output.

Indels, insertion or deletion of nucleotides can happen naturally during the evolution of species or may appear falsely due to sequencing errors and poor sequence quality. Therefore, identifying indels requires the careful study of related sequences. The target sequence was compared with several species' sequences to determine the location of gaps. Two main factors were considered during indel correction: one was that indels should not disrupt the reading frame, as this could result in gene dysfunction, and the other was that stop codon and frame shift should not be present in AA alignments. As the stop codon indicates the termination of the translation process, they cannot be present in the middles of sequences and alignments. They should be always validated and corrected when possible, in the same way that frame shifts were corrected by choosing and modifying the corrected reading frame. A suitable reading frames results in a sequence with no stop codon. The final version of this AA alignment can be seen in Figure 4-25.

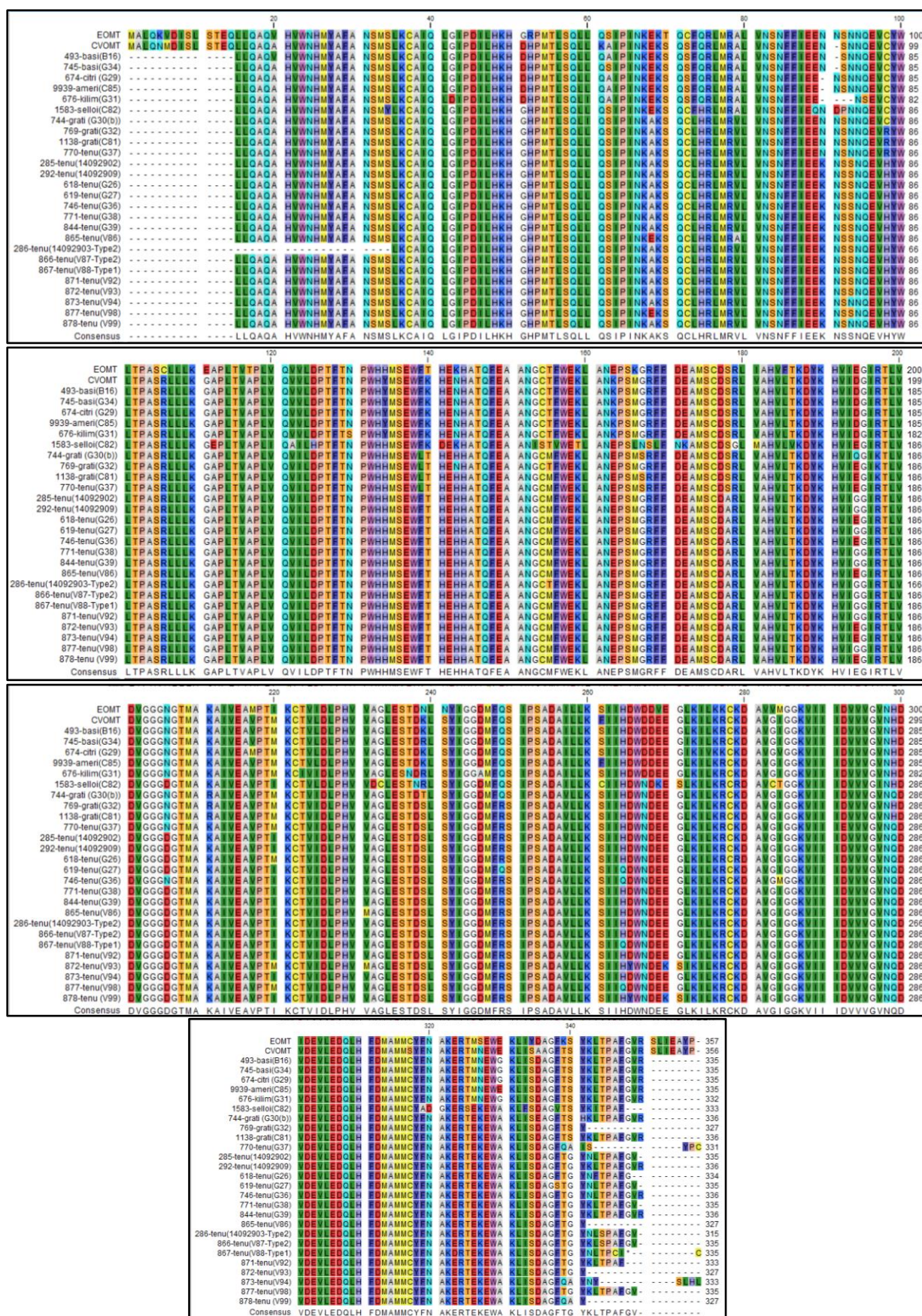


Figure 4-22: A final version of AA alignment with reference sequences [*O. basilicum* EOMT (AF435008) and *O. basilicum* CVOMT (AF435007)], containing around 350 AA of each sample included in the alignment (according to the Rasmol color scheme)

4.3.4.1.1 G37 sample

After careful study the AA alignment results, the G37 sample was found to show resemblances to *O. gratissimum* *EOMT* sequences, even though its identity was confirmed as *O. tenuiflorum*. Consequently, the sample was resequenced using *trnH-psbA* and *EOMT* primer pairs. The next round of results was different from the previous ones: they resolved the problem of both *trnH-psbA* and *EOMT* sequencing showing similarities to *O. tenuiflorum* sequences. The new and old G37 sample sequence alignment highlighted some differences that can be seen in Figure 4-26.

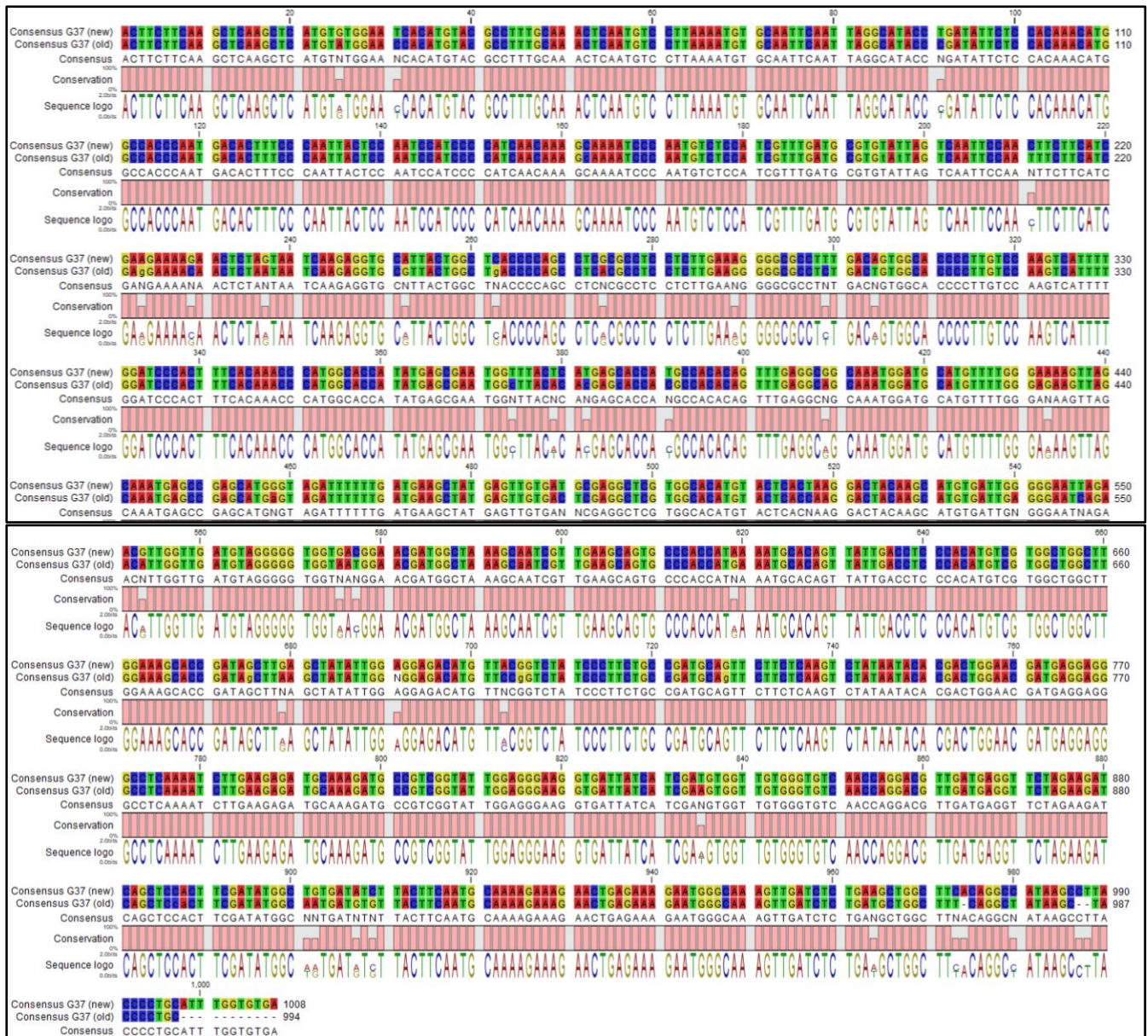


Figure 4-23: An alignment of G37 (old sample mix-up and new pure sample) sequences. The earlier G37 sample showed resemblances to *O. gratissimum* *EOMT* sequence even though it was authenticated as *O. tenuiflorum*. This query was resolved by resequencing the sample using *trnH-psbA* and *EOMT* primers.

4.3.4.2 AA polymorphism study

This alignment study was also helpful in finding AA similarities and differences within *O. tenuiflorum* chemotypes along with other species – *O. gratissimum*, *O. basilicum*, *O. citriodorum* and *O. kilimandscharicum* are included in the alignment. As shown in Table 4-5, AA positions 90, 93, 98, 178, 194, 206, 299, 340 and 342 were highlighted after checking AA variations within different *Ocimum* species.

- Following Anand *et al.* (2014), the highest level of ME was found in young leaves of *O. basilicum*, while the eugenol content was abundant in *O. gratissimum* and *O. tenuiflorum* compared to other *Ocimum* species. Based on these findings, *O. gratissimum*, *O. basilicum* sequences and other *Ocimum* species (including *O. citriodorum*, *O. kilimandscharicum*, *O. selloi*, *O. americanum*) were also included in the study along with *O. tenuiflorum*.
- The alignment is divided into two groups, one in which *Ocimum* species other than *O. tenuiflorum* were included, the other in which all *O. tenuiflorum* chemotypes were included. Some AA positions (90, 98, 178, 261 and 299) were highlighted in Table 4-5, which clearly distinguishes the groups.
- At Position 98, two *O. gratissimum* sequences, G32 and C81, showed the presence of Arginine AA, while all other species including one *O. gratissimum* (G30b) sequence in Group 2 contained Cysteine AA. This finding supports the previous authentication results (Chapter 3) where three types of *O. gratissimum* sequence were included in an alignment study. AA Positions 194 and 299 showed a similar AA polymorphism where one *O. gratissimum* (G30b) sequence stood out compared to the other two (G32 and C81). At Position 194 and 299, Glutamic Acid and Histidine AA were represented by G32 and C81 *O. gratissimum* sequences respectively, while at both positions Glutamine AA was represented by G30b the *O. gratissimum* sequence type.
- At Position 206, all *O. gratissimum* sequences contained Asparagine AA.
- At Position 98, Cysteine AA was present in all Group 2 sequences apart from two *O. gratissimum* (G32 and C81) samples.
- At Positions 178 and 240 a single AA, Serine, was present in all the species in Group 1.

In *O. tenuiflorum*, the levels of methyleugenol were higher than those of eugenol, indicating the presence of the active *EOMT* enzyme. To study the gene structures and relevant protein production (especially the *EOMT* gene in *O. tenuiflorum*), the sequence study of *O. tenuiflorum* chemotypes have been highlighted in Group 2. These highlighted *O. tenuiflorum* specific AAs are the ones that could be responsible for the *EOMT* enzyme being more active in *O. tenuiflorum*.

- At AA Positions 90, 98, 178, 261 and 299 the single AA was present in *O. tenuiflorum* samples.
- At AA Positions 93, 194, 206, 340 and 342 some discrepancies could be seen within *O. tenuiflorum* samples.

Table 4-5: Polymorphic AA positions 90, 93, 98, 178, 194, 206, 261, 299, 340 and 342 are highlighted in AA alignment. *O. basilicum*, *O. citriodorum*, *O. americanum*, *O. kilimandscharicum*, *O. selloi* and *O. gratissimum* were included in Group 1, while all *O. tenuiflorum* chemotypes were included in Group 2. The differences between the two groups were consistent at AA Positions 90, 98, 178, 261 and 299, though some discrepancies can be seen at the remaining AA positions

Species	AA positions on alignment									
	90	93	98	178	194	206	261	299	340	342
<i>EOMT (O. basi)</i>	N	N	C	S	E	N	S	H	S	K
<i>CVOMT (O. basi)</i>	N	N	C	S	D	N	F	H	S	K
<i>O. basi</i> (B16)	N	N	C	S	D	N	S	H	S	K
<i>O. basi</i> (G34)	N	N	C	S	D	N	S	H	S	K
<i>O. citri</i> (G29)	-	N	C	S	D	N	S	H	S	K
<i>O. ameri</i> (C85)	-	N	C	S	D	N	F	H	S	K
<i>O. klima</i> (G31)	-	-	C	S	D	N	S	H	S	K
<i>O. selloi</i> (C82)	N	N	C	S	E	D	C	H	S	K
<i>O. grati</i> (G30b)	N	N	C	S	Q	N	S	Q	S	K
<i>O. grati</i> (G32)	N	N	R	S	E	N	S	H	S	-
<i>O. grati</i> (C81)	N	N	R	S	E	N	S	H	S	K
<i>O. tenu</i> (G37)	K	S	H	A	G	D	S	Q	G	K
<i>O. tenu</i> (14092902)	K	S	H	A	G	D	S	Q	G	N
<i>O. tenu</i> (14092909)	K	S	H	A	G	D	S	Q	G	K
<i>O. tenu</i> (G26)	K	S	H	A	E	D	S	Q	G	N
<i>O. tenu</i> (G27)	K	S	H	A	G	D	S	Q	G	N
<i>O. tenu</i> (G36)	K	S	H	A	E	N	S	Q	G	N
<i>O. tenu</i> (G38)	K	S	H	A	G	D	S	Q	G	K
<i>O. tenu</i> (G39)	K	S	H	A	G	D	S	Q	G	K
<i>O. tenu</i> (V86)	K	S	H	A	E	D	S	Q	G	-
<i>O. tenu</i> (14092903-ME)	K	S	H	A	G	D	S	Q	G	N
<i>O. tenu</i> (V87-ME)	K	S	H	A	G	D	S	Q	G	K
<i>O. tenu</i> (V88-E)	K	S	H	A	G	D	S	Q	G	N
<i>O. tenu</i> (V92)	K	S	H	A	G	D	S	Q	G	K
<i>O. tenu</i> (V93)	K	S	H	A	E	D	S	Q	G	-
<i>O. tenu</i> (V94)	K	N	H	A	E	D	S	Q	A	N
<i>O. tenu</i> (V98)	K	S	H	A	G	D	S	Q	G	K
<i>O. tenu</i> (V99)	K	N	H	A	E	D	S	Q	A	-

Following Gang *et al.*, 2002, there was a single AA difference at Position 261, which was responsible for *EOMT* and *CVOMT* enzymes' substrate discrimination (Gang *et al.*, 2002).

- Based on these results, the AA positions can also be categorised by analysing sequence differences in reference *O. basilicum* positions. As shown in Table 4-6, AA Positions 51, 72, 140 and 163 (along with 261) represent the AA difference within reference *EOMT* and *CVOMT* sequences.
- An experiment set of data shows a mixed picture where all *O. gratissimum* and *O. tenuiflorum* sequences represent similar AA apart from Position 140, where the G30(b) *O. gratissimum* sequence represents *O. basilicum* (*CVOMT*) type AA.
- *O. basilicum* and other *Ocimum* species including *O. americanum*, *O. klimandscharicum*, *O. citriodorum* and *O. selloi* represent similar AA, while *O. tenuiflorum* and *O. gratissimum* all have similar AAs in general.

This observation suggests that all Group 1 sequences are related to the *O. basilicum* (*CVOMT*) reference sequence, while all *O. tenuiflorum* and *O. gratissimum* sequences apart from G30(b) are related to the *O. basilicum* (*EOMT*) reference sequence.

Table 4-6: An example of AA polymorphism at Positions 51, 72, 140, 163 and 261 in AA alignment, where two *O. basilicum* sequences represent different AAs. The differences in the two reference sequences divide the alignment into two categories: one where *O. basilicum* and other *Ocimum* species including *O. americanum*, *O. klimandscharicum*, *O. citriodorum* and *O. selloi* are related to the *O. basilicum* (*CVOMT*) reference sequence, while all *O. tenuiflorum* and *O. gratissimum* apart from G30(b) are related to the *O. basilicum* (*EOMT*) reference sequence

Species	AA positions on alignment				
	51	72	140	163	261
<i>EOMT</i> (<i>O. basi</i>)	G	C	T	E	S
<i>CVOMT</i> (<i>O. basi</i>)	D	S	K	K	F
<i>O. basi</i> (B16)	D	S	K	K	S
<i>O. basi</i> (G34)	D	S	K	K	S
<i>O. citri</i> (G29)	D	S	K	K	S
<i>O. ameri</i> (C85)	D	S	K	K	F
<i>O. klima</i> (G31)	D	S	K	K	S
<i>O. selloi</i> (C82)	G	C	K	E	C
<i>O. grati</i> (G30b)	G	C	T	E	S
<i>O. grati</i> (G32)	G	C	T	E	S
<i>O. grati</i> (C81)	G	C	T	E	S
<i>O. tenu</i> (G37)	G	C	T	E	S
<i>O. tenu</i> (14092902)	G	C	T	E	S
<i>O. tenu</i> (14092909)	G	C	T	E	S
<i>O. tenu</i> (G26)	G	C	T	E	S
<i>O. tenu</i> (G27)	G	C	T	E	S
<i>O. tenu</i> (G36)	G	C	T	E	S
<i>O. tenu</i> (G38)	G	C	T	E	S
<i>O. tenu</i> (G39)	G	C	T	E	S
<i>O. tenu</i> (V86)	G	C	T	E	S
<i>O. tenu</i> (14092903-ME)	G	C	T	E	S
<i>O. tenu</i> (V87-ME)	G	C	T	E	S
<i>O. tenu</i> (V88-E)	G	C	T	E	S
<i>O. tenu</i> (V92)	G	C	T	E	S
<i>O. tenu</i> (V93)	G	C	T	E	S
<i>O. tenu</i> (V94)	G	C	T	E	S
<i>O. tenu</i> (V98)	G	C	T	E	S
<i>O. tenu</i> (V99)	G	C	T	E	S

4.3.4.2.1 *O. tenuiflorum* EOMT sequences

Seven chemotypes (from different geographical habitats of peninsular India) of *O. tenuiflorum* EOMT genomic sequences were submitted to the database (Table 4-3) by Renu and her team (Renu *et al.*, 2014). As all these sequences were identical and contained no variations, they were assembled together to extract one consensus sequence. This single consensus sequence was later aligned with the *O. tenuiflorum* sequences used for this project. This alignment was studied to find similarities and differences between experimental and published *O. tenuiflorum* EOMT sequences.

The consensus sequence seemed to match the majority of experimental sample sequences. Most samples match the consensus sequence, with the exception of a few minor variations.

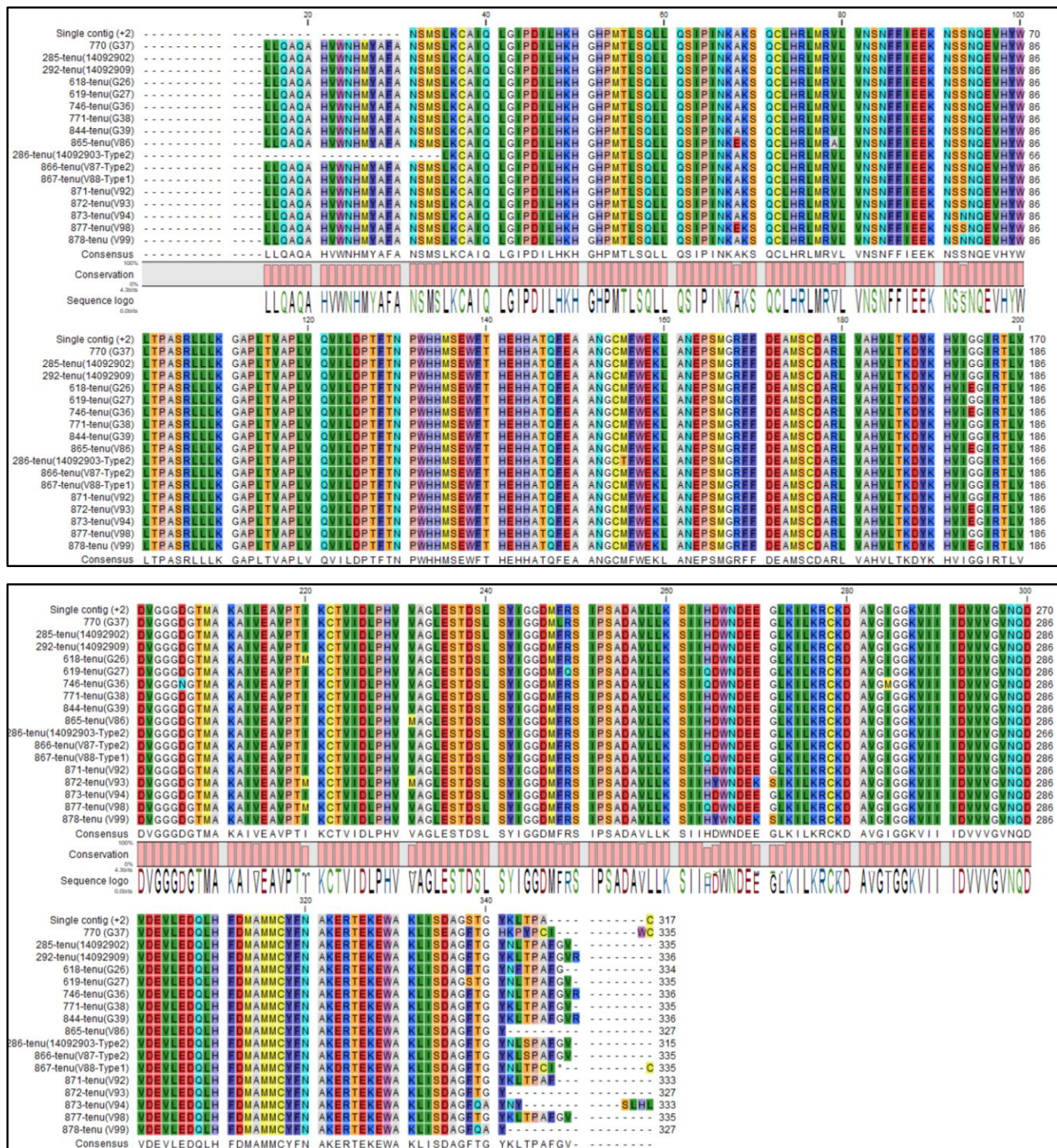


Figure 4-24: *O. tenuiflorum* EOMT AA sequence alignment of the published sequences Ot1: EU622042, Ot2: EU622043, Ot3: EU622044, Ot5: EU622045, Ot6: EU622046 and Ot7: EU622047, represented as a single consensus sequence at the top, and experimental sequences *O. tenuiflorum* (G27), *O. tenuiflorum* (G36), *O. tenuiflorum* (G37), *O. tenuiflorum* (G38), *O. tenuiflorum* (G39), *O. tenuiflorum* (V86), *O. tenuiflorum* (V87), *O. tenuiflorum* (V88), *O. tenuiflorum* (V92), *O. tenuiflorum* (V93), *O. tenuiflorum* (V94), *O. tenuiflorum* (V98), *O. tenuiflorum* (14092902), *O. tenuiflorum* (14092903), *O. tenuiflorum* (14092909), *O. tenuiflorum* (G26), *O. tenuiflorum* (V99). The differences are highlighted with red rectangles.

The published and experimental sequences are similar at the majority of AA positions (Figure 4-27). This alignment proved useful when studying some significant polymorphisms between

an agreed “core” sequence for the *O. tenuiflorum* EOMT sequence and the sequences included in this study. For example, at AA Positions 194 and 264, major differences can be seen between the consensus and experiment sequences, while at AA Positions 68, 79, 93, 155, 194, 206, 220, 231, 265, 270, 271, 284, 338, 339 and 340 the differences are minor (Table 4-7).

Table 4-7: AA polymorphism study between published *O. tenuiflorum* sequences shown as a single consensus sequence in the top row and the experimental *O. tenuiflorum* sequences included in this study

Species			AA positions on alignment														
	Major differences			Minor differences													
AA position	194	264		68	79	93	155	206	220	231	265	270	271	284	338	339	340
<u>Renu</u> consensus sequence	G	H		A	V	S	M	D	I	V	D	E	G	I	S	T	G
<u>O. tenu</u> (G37)	G	H		A	V	S	M	D	I	V	D	E	G	I	F	T	G
<u>O. tenu</u> (14092902)	G	H		A	V	S	M	D	I	V	D	E	G	I	F	T	G
<u>O. tenu</u> (14092909)	G	H		A	V	S	M	D	I	V	D	E	G	I	F	T	G
<u>O. tenu</u> (G38)	G	H		A	V	S	M	D	I	V	D	E	G	I	F	T	G
<u>O. tenu</u> (G39)	G	H		A	V	S	M	D	I	V	D	E	G	I	F	T	G
<u>O. tenu</u> (14092903-ME)	G	H		A	V	S	T	D	I	V	D	E	G	I	F	T	G
<u>O. tenu</u> (V87-ME)	G	H		A	V	S	M	D	I	V	D	E	G	I	F	T	G
<u>O. tenu</u> (V88-E)	G	Q		A	V	S	M	D	I	V	D	E	G	I	F	T	G
<u>O. tenu</u> (G27)	G	Q		A	V	S	M	D	I	V	D	E	G	I	S	T	G
<u>O. tenu</u> (V92)	G	H		A	V	S	M	D	I	V	D	E	G	I	F	T	G
<u>O. tenu</u> (V98)	G	Q		E	V	S	M	D	M	V	D	E	G	I	F	T	G
<u>O. tenu</u> (G26)	E	H		A	V	S	M	D	M	V	D	E	G	I	F	T	G
<u>O. tenu</u> (V93)	E	H		A	V	S	M	D	M	M	Y	K	S	I	F	T	G
<u>O. tenu</u> (V94)	E	H		A	V	N	M	D	I	V	D	E	G	I	F	Q	A
<u>O. tenu</u> (V99)	E	H		A	V	N	M	D	I	V	Y	K	S	I	F	Q	A
<u>O. tenu</u> (G36)	E	Q		A	V	S	M	N	I	V	D	E	G	M	F	T	G
<u>O. tenu</u> (V86)	E	H		E	A	S	M	D	I	M	D	E	G	I	F	T	G

4.3.4.2.2 Eugenol and ME chemotypes within *O. tenuiflorum* samples

The *O. tenuiflorum* AA alignments were carefully studied to identify any AA polymorphisms that correlated with the methyleugenol chemotypes of the sample. As described in Schinle *et al.* (2014) the flavonoids of various *O. tenuiflorum* chemotypes were investigated using the HPTLC technique. All these cultivars were divided into chemotypes, Types 1 and 2, according to their flavonoid profile. According to test results, those cultivars with a Type 1 flavonoid profile were also categorised as a high methyleugenol type, while the Type 2 flavonoid chemotypes were identified as having a high eugenol type. An AA alignment shown in Figure 4-28 contains sequences from two Type 2 chemotype samples (286 (14092903) and 866 (V87) and a sequence from one Type 1 sample (867 (V88) from CAMAG. There is no consistent difference (apart from AA Position 141, where one Type 2 sequence (Sample 286) is different from other two) between the chemotypes, but with the small sample size, it was difficult to say this with certainty.

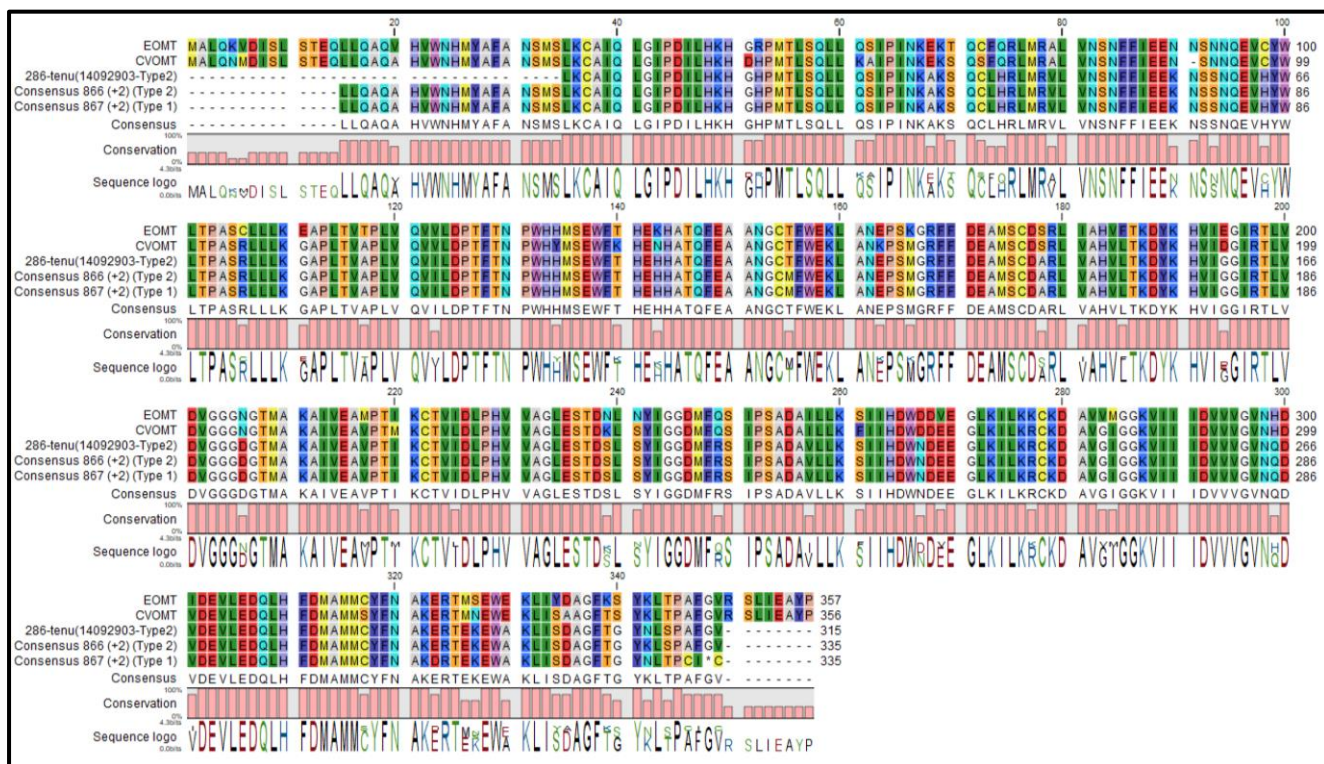


Figure 4-25: An AA alignment contains two Type 2 flavonoids (286 (14092903) and 866 (V87) sample sequences) and one Type 1 flavonoid (867 (V88) sequence) from the CAMAG study. There are no consistent differences between these samples apart from the one at AA Position 155

In summary, several other comparisons can be made from the AA alignment of EOMT:

1. Positions that distinguish *O. tenuiflorum* from all the other species are AA position No 139, 167, 178 and 299.
2. Positions that vary within *O. tenuiflorum* correlated with their chemotype. Examples are AA Position 155.
3. Positions that distinguish the *O. basilicum* cluster including *O. basilicum*, *O. americanum*, *O. citriodorum* and *O. kilimandscharicum*, *O. selloi* from *O. tenuiflorum* and *O. gratissimum*. Examples are AA Positions 68, 79 and 98.
4. The differences in *O. basilicum* EOMT and *O. basilicum* CVOMT reference sequence positions can be studied in relation to other *Ocimum* species. Examples are AA Positions 51, 72, 134 and 194.

4.3.5 Phylogenetic analysis

Various algorithms including UPGMA (the Unweighted Pair Group Method with Arithmetic mean), NJ (Neighbor Joining), FM (Fitch-Margoliash) and ME (Minimum Evolution) are

available online which are useful construct a phylogenetic tree. UPGMA is a simple and fast method yet categorises organism/species based on overall similarity regardless of their evolutionary relationships. In NJ each pair is evaluated, and the sum of all branch lengths is calculated and arranged as a tree. It can, however, only produce a tree by avoiding other possibilities, and as a result a biased tree is prepared. Both FM and ME observe pair-wise distances to the expected distance of a tree. ME has first to calculate external nodes and then use it to calculate internal ones. ME therefore stands out among all distance-based algorithm methods, yet it is slower than NJ. Web services such as ‘*Phylogeny.fr*’ that are dedicated to reconstructing and analysing phylogenetic relationships, and visualisation apps like Jalview, are also available online to prepare phylogenetic trees.

Nucleotide sequences of desired species and deduced AA sequences were used to study site by site differences in the tree. Various *Ocimum* species’ AA sequences were used to prepare the phylogenetic tree as shown in Figure 4-29 by using ‘*Phylogeny.fr*’ platform. ‘*Phylogeny analysis*’ tab contains the range of bioinformatics programs including ‘*One click*’, ‘*Advance*’ and ‘*A la carte*’ options are within ‘*Phylogeny analysis*’ tab. Of these, the ‘*A la carte*’ mode was used to prepare the tree by using experimental data. As shown in Figure 4-29, there is a major branch between the *O. basilicum* group and *O. tenuiflorum* along with *O. gratissimum*. There is then a sub-branch between *O. tenuiflorum* and *O. gratissimum*. By observing this tree, these *Ocimum* species can be divided into three groups. Although each of them coming from common ancestor, over the time they develop some interesting characteristics and therefore some interesting arrangements can be seen in Figure 4-29. All *O. gratissimum* species sit in first group suggesting some unique features evolved over generations. All *O. tenuiflorum* species sits in second group. Other species including *O. americanum*, *O. basilicum*, *O. citriodorum* and *O. klimandscharicum* made third group showing some close relationships with *O. basilicum* (CVOMT) and *O. basilicum* (EOMT) reference sequences, among which *O. americanum* shows connection with *O. basilicum* (CVOMT), while *O. selloi* shows long distant relationship with all other *Ocimum* species. G34 and B16 *O. basilicum* sequences could be the direct and non-direct descendant of *O. basilicum* (EOMT).

The tree mainly shows the diversity within the *O. tenuiflorum* samples, with the majority being very similar, but with two or three minor sub-groups.

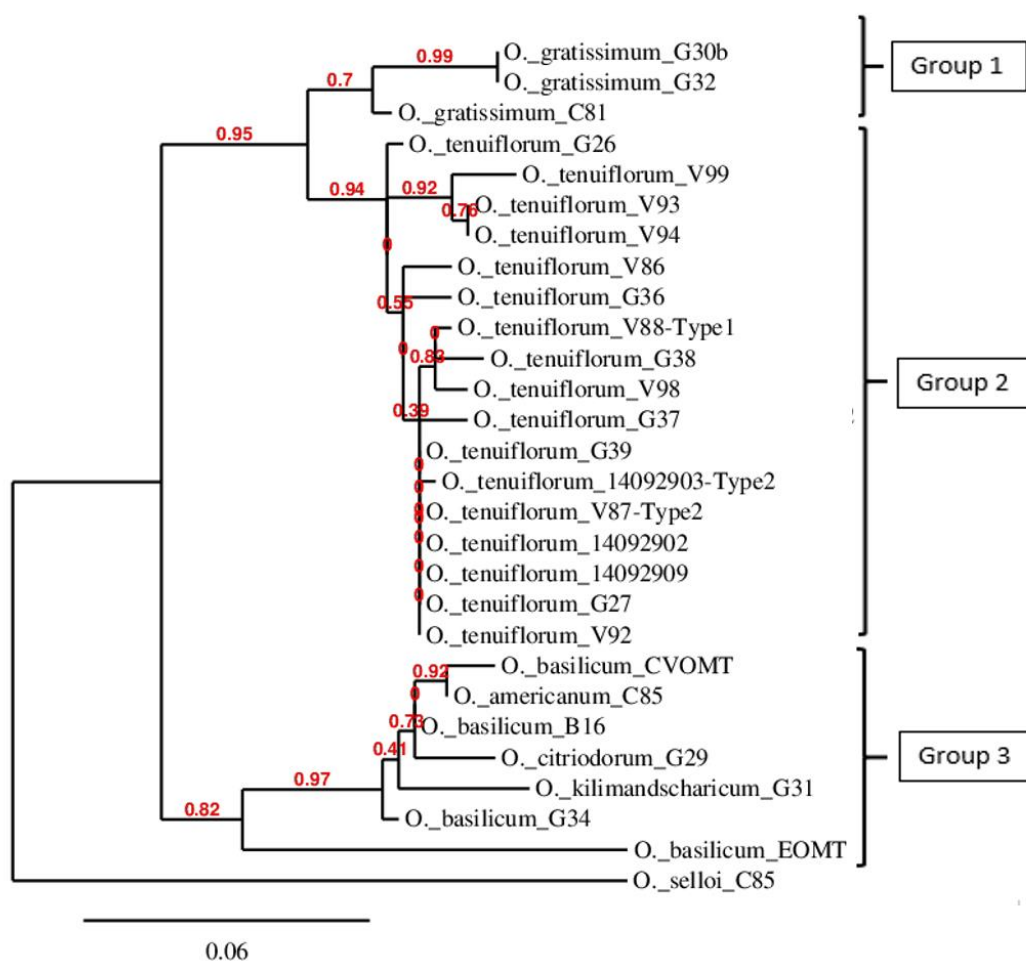


Figure 4-26: Phylogenetic tree construction using Phylogeny analysis tool with 'A la Carte' mode in the Phylogeny.fr program (with default settings). Phylogenetic relationship of *Ocimum* species. The length of the AA is around 350. AF435007 (*O. basilicum* CVOMT) and AF435008 (*O. basilicum* EOMT) (Gang *et al.*, 2002) were used as reference sequences (highlighted in the figure) to categorise other *Ocimum* species into relevant groups. Branch lengths are based on the number of amino acid substitutions per site between the sequences included in this study. Clusters of sequences showing very strong relationships are represented as Groups 1, 2 and 3. Bootstrap values are shown below each node in %. The AA sequences included in this analysis are as follows: *O. tenuiflorum* (G27), *O. tenuiflorum* (G36), *O. tenuiflorum* (G37), *O. tenuiflorum* (G38), *O. tenuiflorum* (G39), *O. tenuiflorum* (V86), *O. tenuiflorum* (V87), *O. tenuiflorum* (V88), *O. tenuiflorum* (V92), *O. tenuiflorum* (V93), *O. tenuiflorum* (V94), *O. tenuiflorum* (V98), *O. tenuiflorum* (14092902), *O. tenuiflorum* (14092903), *O. tenuiflorum* (14092909), *O. tenuiflorum* (G26), *O. tenuiflorum* (V99), *O. gratissimum* (G32b), *O. gratissimum* (C81), *O. gratissimum* (G32), *O. basilicum* (B16), *O. basilicum* (G34), *O. citriodorum* (G29), *O. kilimandscharicum* (G31), *O. americanum* (C85) and *O. selloi* (C82).

4.4 Discussion

4.4.1 Obtaining the consensus sequences

In genetics, studying the target gene has mostly involved its sequence analysis to determine the gene's role in the relevant organism. This project has therefore evolved around the targeted *OMT* gene study, which began by isolating and analysing DNA sequences. DNA samples were first obtained from various *Ocimum* species by using the techniques of

molecular biology, then sequence study used a variety of bioinformatics tools. This is another approach to the study of relevant genotypes also being used to determine *OMT* gene functions. This perspective provided detailed information regarding the *OMT* gene structure and function in various *Ocimum* species.

This study includes several cDNA sequences representing potential *OMTs* with varying degrees of similarity to previously published cDNA sequences of *O. basilicum* (Gang *et al.*, 2002). With the help of 90 per cent identical reference sequences (*O. basilicum EOMT* and *O. basilicum CVOMT*), evolutionary relationships to a large family of plant *OMTs* was studied in detail.

4.4.1.1 DNA quality

Several precautions were taken while extracting and handling the DNA samples during this study. Each sample's concentration and purity level was measured using a nanodrop instrument. The quality of some samples was, however, poor compared to others. *O. selloi* (C82) was one of those whose quality scored comparatively low. As a result, while assembling their forward and reverse reads, several conflicts appeared; these were carefully resolved before using their consensus DNA sequences for alignment purposes.

The search for similar regions among different biological sequences was made reliable with the help of good-quality sequences. Sequences such as *O. selloi* were not up to standard, however, and it is debatable whether the major differences within alignments are trustworthy. However, by aligning poor quality sequences along with reliable *O. basilicum* reference sequences and other good-quality experimental datasets, the sub-optimal alignment results can be mitigated.

4.4.1.2 Various *EOMT* PCR primers

As shown in Table 4-1, various primer pairs were constructed in order to obtain the full-length *OMT* gene. Some primer sequences were obtained from previous literature (Gang *et al.*, 2001; Gang *et al.*, 2002), among which the *EOMT* primer sequences used in Renu *et al.* (2014) proved useful in successfully identifying the *OMT* region. Apart from using published primer sequences, a number of primers were constructed and used after studying the *EOMT* gene.

Among all these primer pairs, the *EOMT* primer sequences from Renu *et al.* (2014) were found within the GOI sequence as shown in Figure 4-5, although the short amplicons of *EOMT* primers proved unable to efficiently cover the internal region of the *OMT* sequence. The internal primer pair was designed in order to overcome this difficulty. The combination of these two-primer pair sequences (*EOMT* and internal) were used to build a library of all *EOMT* gene consensus sequences. After several modifications in these DNA consensus sequences, an AA alignment was prepared in order to study polymorphism within *Ocimum* sequences.

4.4.2 Full-length *O-methyltransferase* gene study

The main aim of this research was to compare the *O-methyltransferase* gene sequences in closely related *Ocimum* species. This has been successfully achieved by studying diverse *Ocimum* species including *O. tenuiflorum*, *O. gratissimum*, *O. basilicum*, *O. americanum*, *O. kilimandscharicum* and *O. africanum*. Along with studying *OMT* gene in various *Ocimum* species, this perspective has also been used to study the same gene structure in different chemotypes of *O. tenuiflorum*.

The completion of this project depended entirely on the efficient primers suitable for the target gene type. The success of primer design depends mainly on the use of suitable software, successfully achieved for this project by using a variety of software including IDT primer design. The primer designs for suitable gene types yielded high-quality results and proved to be species-specific within the samples tested. They all managed to produce the desired size fragments for the majority of *Ocimum* species types. The details of the published DNA sequences play an important role in this type of investigation, as the efficiency of specially designed primers depends entirely on the available sequence data types. Due to the availability of advanced sequencing facilities, all DNA data will in future be accessible for all medicinal and economically important plants which can become pioneers in designing this technique.

The primer design is an experimental test that requires DNA as a template. This was used in this project in various forms including DNA extracted from fresh plant samples and vouchered DNA samples. These results were further tested against a variety of DNA templates. Designing and testing different primer pairs and setting suitable thermocycle

conditions was a labourious task. Even though the latest primer designing software is efficient enough to design effective primer pairs, it can take a long time to settle on the first protocol in the laboratory. Introducing primer pairs one by one, good-quality results can be achieved by trial and error, where optimisation is required to remove unwanted effects or components from the experiment. Once procedures are established, the final protocol could be accomplished through step by step guidance.

Renu's *EOMT*s was the first set of primers used for this study. The primer sequence and thermocycler programme details were taken from Renu *et al.* (2014). The various *Ocimum* species included in this study gave positive results with these primer pairs, making it the most reliable primer pair for this study. They were therefore the first choice of primer pairs when any new protocols were introduced in this study. These primer pairs were able to sequence the full-length amplicon, but not in sufficient quality along its whole length to guarantee confidence in the contig. One more primer pair was thus designed to cover the gene's internal region, as this region was not efficiently covered by Renu's *OMT* primers.

The second primer pair called Internal primers (Int-F and Int-R) were designed to provide sequence data covering each half of the gene with an overlap in the middle, so that the gene sequence comprised good-quality forward and reverse reads along its length. As shown in Figure 4-5, they sit neatly in the middle region and cover all internal regions not efficiently covered by the *EOMT* primers. This 'Nested PCR strategy' is a modification of PCR that was designed to improve sensitivity and specificity. Here, Nested PCR involves the use of two primer sets and two successive PCR reactions in the form of *EOMT* and internal primer pairs.

By designing internal primers, the entire *OMT* internal region has been successfully covered, while most of the external region is covered by *EOMT* primers. The initial and end regions were, however, still not covered by any primer pair. This was absolutely essential in preparing full-length *OMT* cDNA sequences for cloning purposes. To fulfil this requirement a third primer pair, *OMT*-1, which includes the two forward primers *EOMT*-1 and *CVOMT*-1, as well as *OMT*-1074 primer pairs, were designed. The details and relevant results can be found in the next chapter.

4.4.3 Phylogenetic analysis of OMT gene within various *Ocimum* species

The phylogenetic tree was constructed by using all AA sequences included in the alignment, along with the two *O. basilicum* reference sequences *EOMT* and *CVOMT*, which represents the hypothetical evolutionary relationships among a group of *Ocimum* species. All AA sequences were compared between species and used to build phylogenetic trees. The closely related species including all samples within *O. tenuiflorum* and *O. gratissimum* showed few differences in sequences, while other less-related species contain more sequence differences. In other words, there is a common ancestor for all *O. tenuiflorum* and *O. gratissimum* species as they all arise from the single branch showing 97 per cent bootstrap value (is used to estimate the confidence of the branches in a phylogenetic tree). The same can be seen for the remaining *Ocimum* species, which are at the branch point with the common ancestor of the *O. basilicum* (*EOMT*) sequence. Interestingly, one of the branches represents the *O. basilicum* (*CVOMT*) sequence, which supports the hypothesis that both *EOMT* and *CVOMT* sequences evolved from the same gene.

The line leading to the species *O. selloi* represents the species' ancestors since it diverged from other species in the tree. The genetic distance between all other *Ocimum* species and *O. selloi* is highest, as the branch's distance from the *O. basilicum* (*EOMT*) sequence is greater than any other.

The phylogenetic tree construction is based on the analysis intended to detect intra- and inter-specific genetic diversity, and on the study of the phylogenetic relationship in the genus *Ocimum*. A tree has been divided into different groups in order to determine the genetic base and phylogenetic status of the *Ocimum* species. All *O. gratissimum* species form a first group, all *O. tenuiflorum* sequences form the second, with the variety in percentages being similar between the two. More than half of the sequences in these groups have a bootstrap value of around 50 per cent. All other *Ocimum* species display a similarity with the *O. basilicum* (*CVOMT*) reference sequence, while *O. selloi* stood out as different from all other *Ocimum* sequences. *O. basilicum* (G34) and *O. basilicum* (B16) sequences have higher and lower bootstrap values respectively than the *O. basilicum* (*CVOMT*) reference sequence. It must be born in mind that predicting results on the basis of bootstrap values is not reliable, as all genes are continuously undergoing evolution. High bootstrap values supporting a phylogenetic tree structure, although researchers favour discussing low bootstrap values to

reinforce the claim that a particular taxon is not well supported by a particular set of data (Wiesemüller and Rothe, 2006).

Gang *et al.*, 2002 suggest that there are two genes in *O. basilicum* – the *EOMT* and *CVOMT* types. The experimental data show that the genes in related species such as *O. citriodorum* and *O. kilimandscharicum* code for the *EOMT* gene (as per the AA alignment, Serine AA is at Position 261 as described in Table 4-6), while *O. americanum* has Phenylalanine AA at Position 261, as per the *O. basilicum* (*CVOMT*) sequence. Phylogenetic analysis also supported this set of data.

4.4.4 The key discoveries of sequence analysis

It was observed in the literature study that very few research findings are available regarding the role of target secondary metabolites (eugenol, chavicol and their methyl derivatives) in *Ocimum*, and some of their metabolite pathway steps have still not been determined. Gang *et al.* (2002) also identify two different enzymes with some shared characteristics, *EOMT* and *CVOMT*, which are responsible for this methylation process, as well as isolating related genes. They also mention the ratios of these two compounds, which are dissimilar in different *Ocimum* types. Moreover, young and old leaves contain both high and low levels of these compounds. Based on these findings, different *Ocimum* species were collected for this project in order to analyse their sequences to discover the factors affecting the production of secondary metabolites in each Tulsi type. As these two enzymes are similar to each other, their chemical separation is difficult (Wang, 1999), which leads the present research towards a molecular analysis of these enzymes. They are mainly produced in the glandular trichomes present on leaves and the stem of *Ocimum* plants. There is the possibility that two polypeptides are involved in the methylation of eugenol and chavicol. The study of these polypeptide sequences is the main means by which to identify the activity of these enzymes. Besides analysing their sequences, they were further cloned in *E. coli* bacteria to study their protein production levels.

The summary of the sequence analysis and phylogenetic data divided all sequences into two groups, one including all *O. tenuiflorum* sequences, the other including sequences representing different *Ocimum* species (*O. gratissimum*, *O. selloi*, *O. citriodorum*, *O. kilimandscharicum*, *O. americanum* and *O. basilicum*). This result data has revealed the

possibility that different amino acid sequences could be responsible for different enzyme activities in the different *Ocimum* species.

4.4.5 Combination of sequence analysis and phylogenetic study

A number of characteristic features distinguishing *Ocimum* species can be identified by combining the sequence alignment findings and phylogenetic analysis. The distantly related *O. selloi* position on the phylogenetic tree and the sequencing results are compatible. Many discrepancies within AA arrangements were observed while comparing the *O. selloi* sequence with others and the phylogenetic tree also related it distantly to other *Ocimum* species. A total of four *O. basilicum* species were included in the alignment, where *O. basilicum* (*EOMT*) and *O. basilicum* (*CVOMT*) were reference sequences (Gang *et al.*, 2002). Both experimental *O. basilicum* species, B16 and G34, were identical to the reference *CVOMT* sequence, apart from AA position 261. This confirms their position on the phylogenetic tree, where they resemble the *O. basilicum CVOMT* sequence. The finding of this research is different from Gang *et al.*, 2002, as there was no *O. basilicum* sequence representing the *EOMT* gene; rather, both B16 and G34 *O. basilicum* sequences resembled the *O. basilicum CVOMT* gene. *O. americanum* was the only species sequence that represented phenylalanine AA at position 261, and therefore it could reliably be identified as a *CVOMT* gene.

4.5 Conclusion

As described earlier, the peltate glands of basil leaves are major producers of phenylpropanoids. The steps in the biosynthetic pathway that produce these phenylpropanoids have not been thoroughly investigated (Gang *et al.*, 2001). Therefore, two different cDNA sequences of *EOMT* and *CVOMT* were earlier isolated from different basil lines. These two cDNAs were expressed in bacteria to produce the relevant proteins. The difference of one AA (*EOMT* - S261 and *CVOMT* - F260) was also confirmed through site-directed mutagenesis (Gang *et al.*, 2002). *EOMT* genomic and cDNA sequence activities were subsequently investigated in *O. tenuiflorum* chemotypes (Renu *et al.*, 2016). Recent biochemical studies have demonstrated how these enzymes can modulate the levels of phenylpropanoids in *O. gratissimum*, *O. basilicum* and *O. tenuiflorum* species (Anand *et al.*, 2016).

In this way, all previous studies have only included certain *Ocimum* varieties, while the present research involves several *Ocimum* species and their varieties. The novel finding of the experimental dataset is the identification of numerous *Ocimum* cDNA sequences at the same time by preparing meaningful alignment result outputs.

The results presented in this chapter demonstrate the characterisation of phenylpropanoid biosynthetic enzymes, *EOMT* and *CVOMT*, through a range of cDNA sequences. A detailed analysis of their AA sequences represented in the alignment suggest that the *Ocimum* species included in this experiment could be capable of synthesising the target phenylpropenes, and that their sequences play an important role to study related pathways. Around two thirds of the sequences from the alignment were similar to the *EOMT* sequence, while the remainder was likely to be represented the *CVOMT* sequence. The similarity between *O. tenuiflorum* and *O. gratissimum* raises the possibility that one prominent *EOMT* gene is present in all *Ocimum* sequences, and that is converted into a different enzyme during the process of evolution or due to some other factor, or that the same *EOMT* enzyme behaves differently in the presence of different substrates. In other words, the way *EOMT* converts eugenol into methyl eugenol could mean that it also converts chavicol into estragole, though possibly not as efficiently as *CVOMT*. The *CVOMT* enzyme is also present in the *Ocimum* species, again possibly with the exception of *O. tenuiflorum* and *O. gratissimum*.

4.5.1 Further work

AA similarities and differences studied during this project could be correlated to high and low contents of eugenol, ME, chavicol and estragole in the respective species and their chemotypes. This remarkable AA polymorphism details could be further used to study extensive transcriptional reprogramming of *EOMT* and *CVOMT* enzymes at pre-flowering and post-flowering stages.

5 Acquisition and cloning of full length cDNAs representing *OMT* genes from different *Ocimum* species

5.1 Introduction

5.1.1 *O-methyltransferase* enzymes (comparisons of *EOMT* and *CVOMT* enzymes)

O-methyltransferase enzymes protect plants against pathogen attack by inducing lignin synthesis (Pichersky and Gang, 2000). This group of enzymes methylated several secondary metabolites including flavonoids and phenylpropanoids (Gang *et al.*, 2001). Among them, eugenol, chavicol and their methyl derivatives are present in essential oils of the Lamiaceae family plants (Lange *et al.*, 2000; Jirovetz *et al.*, 2003 and Viña and Murillo, 2003). They use *S-adenosyl-L-methionine* as the methyl donor during the methylation process. Both *EOMT* and *CVOMT* enzymes share striking similarities. In *O. basilicum*, genomic DNAs have 843 bp ORFs and 90 bp long introns, both genes' expression level peaks in young leaves and is reduced in mature ones. Their 40 kDA proteins are generally stable below 36°C (Wang and Pichersky, 1998; Renu *et al.*, 2014). The activation of one enzyme results in a high level of its methyl derivatives. In other words, an expression of the *EOMT* enzyme is responsible for methyl eugenol while the *CVOMT* enzyme is responsible for estragole production. This finding indicates an activity of two different *O-methyltransferase* enzymes, though it depends on substrate availability and specificity within different *Ocimum* species and their relevant chemotypes (Gang *et al.*, 2002).

Eugenol and chavicol compounds are methylated by two different though closely related enzymes, *EOMT* and *CVOMT*. A single AA difference (Ser-261 in *EOMT* and Phe-260 in *CVOMT*) has appeared to be responsible for these enzymes' substrate preference. Gang *et al.* (2002) indicate that the ratio of *EOMT* and *CVOMT* enzyme activities were not same between different sweet basil varieties. It is not clear, however, whether *CVOMT* has evolved from *EOMT* or vice versa. To determine this requires further analysis of these two enzymes within different *Ocimum* species.

5.1.2 Overview of DNA cloning

Researchers can use bacteria to make multiple copies of a Gene of Interest (GOI). The selected gene is inserted into a plasmid by a ligation process with or without the use of restriction enzymes (REs). This plasmid is then inserted into a host-like bacteria where many copies of its GOI are produced. This process is called DNA cloning (Figure 5-1). With the help of selection methods (mainly antibiotic resistance assays), the bacteria colonies are screened to check the presence of the GOI (Reece *et al.*, 2011). There are several techniques available to screen colonies (see Section 5.2.1.6 for more details). Once a gene's presence in a colony has been confirmed, a larger culture is produced to make the desired protein or plasmids. Colony screening requires special attention, because when a GOI is cut and pasted into a plasmid using RE, it can be inserted either into the right or wrong orientation. The plasmid sometimes closes without taking the GOI, too.

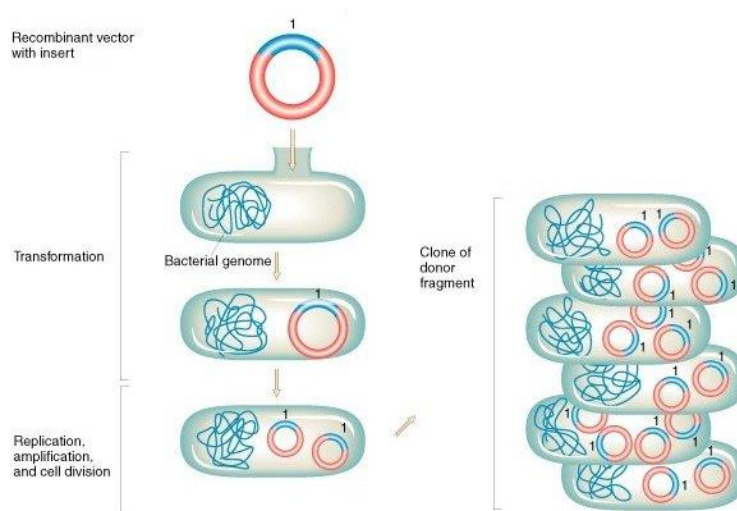


Figure 5-1: Molecular cloning of recombinant DNA in *E. coli* (Ref.: <https://www.mun.ca/biology/scarr/Gr12-05.html>)

DNA cloning can be divided into three main steps:

1. Preparing DNA fragments

A DNA fragment has been digested by suitable RE to add short, single-stranded overhangs at each end. This same overhang sequence has been added to the desired plasmid. Later, when two molecules with the same overhangs come together, they can be joined by using the DNA ligase.

2. Transformation and antibiotic selection

During the transformation process bacteria cells are given a shock through heat shock treatment or electroporation to encourage them to accept foreign DNA. Once the plasmid

with a GOI is inserted into a bacterium, it can be recognised through the antibiotic selection process. As the plasmid contains the selected antibiotic resistance gene, the bacteria that took up the plasmid can survive in the presence of this antibiotic while others without the plasmid die. At the end of the process, multiple colonies containing the plasmid with the GOI are produced.

3. Protein production

Once the colony with the correct GOI has been identified, the particular colony with the plasmid can be grown in large cultures. When the desired quantity of culture has been grown, the target protein can be produced by giving the bacteria a termination signal using relevant chemical.

5.1.3 *Escherichia coli* (*E. coli*) as a model system

Baculovirus and yeast expression systems are hugely popular due to the close resemblance between the processing and transportation properties of expressed proteins to that of mammalian cells (Ikeda *et al.*, 2006; Renu *et al.*, 2014). Post-translational modification involves some fundamental steps including the folding and regulating of proteins, their targeting towards relevant subcellular compartments and the regulation of catalytic activity if protein acts as an enzyme or if managing the protein signalling function is associated with signal transduction pathways (Gomord and Faye, 2004). Of all hosts, *E. coli* seems to be the most convenient and easily available bacterial host system that is useful for expressing the majority of plant proteins (Hartley, 2006). An immense catalogue of expression plasmids, many cultivation strategies and readily available engineered strains makes this organism the most popular expression platform for the high-level production of heterologous proteins. High yields of many plant enzymes (Carvajal-Vallejos *et al.*, 2007), some of the plant membrane proteins (Uozumi *et al.*, 2000) and small plant proteins with a fusion partner (Murphy *et al.*, 1997) can be obtained by using the *E. coli* system. In short, *E. coli* is an ideal expression system, with well-established protocols of the majority of recombinant proteins. Protein production has been affected by several factors, mainly by host type.

5.1.4 Cloning technologies for protein expression

Researchers use a variety of tools and techniques that help isolate ranges of proteins to design cloning and expression protocols. As described in the field of cell biology, an understanding

of proteins is linked with the study of gene expression. The three-dimensional structure of proteins has been determined by the sequence of AA. The study of protein structure helps to understand their action. It links to the gene expressed in a particular cell which determines its function. The available literature and online sharing tools enable scientists to overcome difficulties and produce proteins on a large scale. However, the selection of a suitable expression system that can produce the right size of protein at such a scale is a challenging task. The ideal expression system should produce a large number of proteins, grow quickly and incorporate easy protocols. Bacteria, insects, yeasts, plants and mammals comprise distinct expression systems when used for research purposes (Saluta and Bell, 1998).

1. Yeast cells

Even though prokaryotic systems are widely used as an expression platform, the yeast system is a powerful means of producing certain proteins. Yet yeast demands the allocation of much more time and effort than bacteria. The final product produced by the yeast system is different to bacteria, as it has capability to secrete, process and modify proteins.

2. Insect/Baculovirus cells

The protein produce using this cell line is highly expressed and contains disulphide bands that are efficiently generated. It can provide the majority of post-translation modifications found in mammalian cells.

3. Mammalian cells

The mammalian cell expression system can recognise the signals for secretion, processing and synthesis of the majority of eukaryotic proteins. Post-translational modifications and protein folding are the main characteristics of the recombinant proteins produced by the mammalian expression system. These characteristics are important to complete each biological activity. Most of the pharmaceutical industry's products are therefore produced through the mammalian expression system.

4. Cell-Free Protein Expression

This expression method uses no living organisms to produce protein. Rather, it relies on the readymade biological machinery in a cell-free system. The main advantage of using

this system is that there is no need to maintain the cell's viability, though toxins can be produced.

A detailed knowledge of targeted proteins is fundamental to designing cloning strategies. The proteins must be purified from relevant recombinant sources. Protein purification is a laborious task compared to designing genome sequences. There are several difficulties involved, including designing the right gene sequence, inserting it into the host in the correct orientation, using appropriate inducers and synthesising the amounts of proteins needed for careful implementation to obtain the desired recombinant protein. Scientists therefore prefer to work in parallel by designing multiple protocols at the same time (Hartley, 2006).

5.1.4.1 *E. coli* expression systems

IPTG (Isopropyl β -D-1-thiogalactopyranoside) is a popular induction agent in molecular biology. This compound is a molecular mimic of allolactose, a lactose metabolite that triggers transcription of the lac operon, and it is therefore used to induce protein expression where the gene is under the control of the lac operator and hence is used to induce protein expression in *E. coli*. This is the most commonly used technique in protein expression assays to produce recombinant proteins in bacteria. A modified lac operon system has been engineered in commercial *E. coli* system with the gene coding for T7 RNA polymerase. During IPTG induction this lac repressor protein becomes inactive, allowing T7 RNA polymerase to initiate the process of transcription and translation. In some *E. coli* expression systems, the gene coding for T7 lysozyme has also fallen out, which reduces background expression levels prior to protein induction. The IPTG-induced gene expression system is illustrated in Figure 5-2.

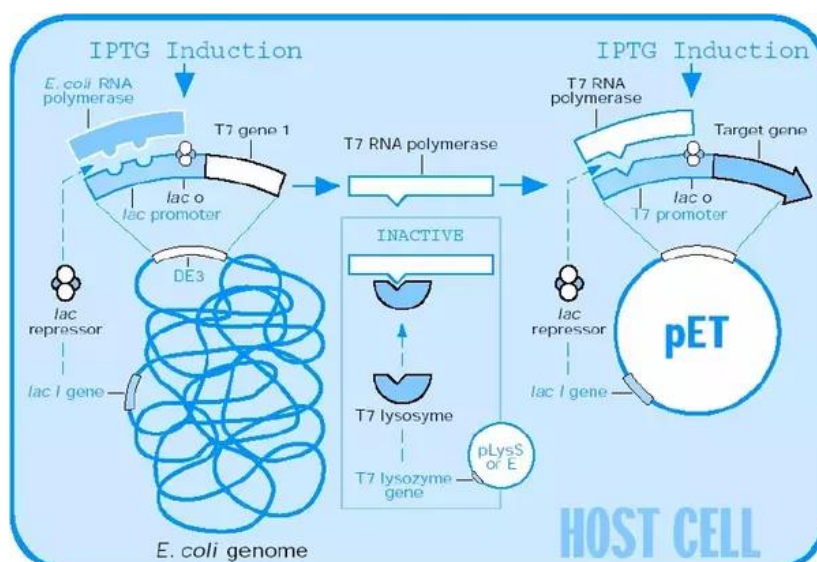


Figure 5-2: Illustration of IPTG-induced gene expression (Ref.: Novagen pET expression manual)

5.1.5 The aim of the chapter

This chapter initially aims to obtain full length *OMT* cDNA sequences from the *O. tenuiflorum*, *O. gratissimum* and *O. basilicum* species. Later, these sequences representing *OMT* genes will be successfully cloned in *E. coli* to study the *OMT* gene from different *Ocimum* species.

5.2 Materials and methods

The material, procedure and methods described in this section are specific to this chapter only. General material and method details can be found in Chapter 2. Additional details can be found in Appendix at the end of this document.

5.2.1 pGEM vector cloning

The *pGEM®-T Easy* vector (Catalog No. - A1360, Promega) strain was used to carry out an initial part of cloning strategies (See figure 5-3).

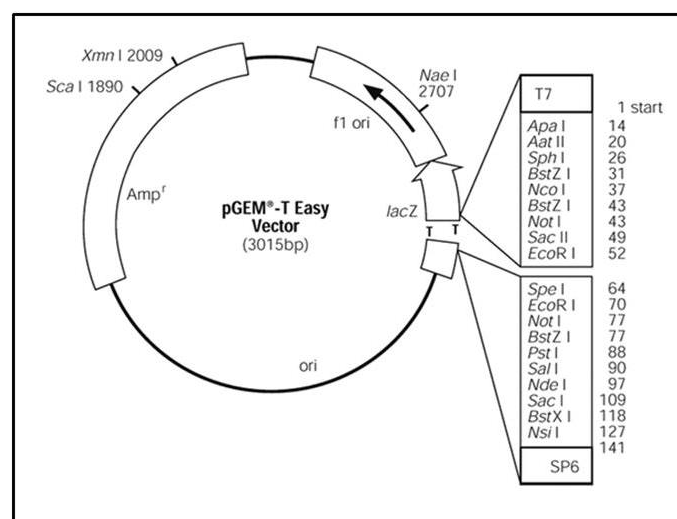


Figure 5-3: pGEM®-T Easy vector map

5.2.1.1 Fresh *Ocimum* samples

For RNA isolation, young leaves were harvested from authenticated *Ocimum* plants growing in De Montfort University's greenhouse. RNeasy Plant Mini Kit (Qiegen, 2012) was used to extract RNA, the Thermo Scientific - TURBO DNA-free™ Kit (Catalog number: AM1907) was used to remove traces of genomic DNA and the Bioline SensiFAST cDNA synthesis kit (Cat. No. BIO-65053) was used to synthesise the cDNA.

5.2.1.2 RNA extraction and cDNA synthesis gene-specific primer design

Total RNA was extracted from young *Ocimum* leaves and cDNA was synthesised (as described in Chapter 2) using gene-specific primers. To design these primers, nucleotide sequences were aligned to identify conserved regions of related sequences. Three primers were designed to cover the *EOMT* and *CVOMT* gene full sequences (Table 5-1).

Table 5-1: *EOMT*-F-1 and *CVOMT*-F-1; *OMT*-R-1074 primers.

<i>EOMT</i> -F-1	5'-ATGGCATTGCAAAAAGTAGA-3'
<i>CVOMT</i> -F-1	5'-ATGGCATTGCAAAATATGGA-3'
<i>OMT</i> -R-1074	5'-TTAAGGATAAGCCTCTATGAG-3'

The thermocycler programme conditions for cDNA synthesis were 25°C for 10 minutes, 42°C for 15, 48°C for 15, 85°C for 5 and 4°C hold. All cDNA products were visualised using 2% w/v agarose gel.

5.2.1.3 Cloning strategy

The *E. coli* (Cambio Ltd., Cat. No. EC10005) strain has been used to prepare electrocompetent cells by using the protocol described in following section.

5.2.1.3.1 Preparation of Electrocompetent *E. coli*

The fresh colonies were grown from existing strains using LB plates. One of these colonies was selected to inoculate 5ml of liquid LB in a 15ml tube. The tube was placed on a shaking incubator at 37°C overnight at a speed of 220rpm. Next day the bacteria cells were collected by centrifuging at 3500rpm for 10 minutes. Supernatant was discarded, and the pellet was resuspended in 1ml of ice-cold water three times, then in 10 per cent glycerol (Cat. No. - G5516 Sigma Glycerol) wash. The pellet was stored in approximately 100µl of glycerol and stored at -80°C.

5.2.1.4 Ligation and cloning of cDNA fragments to *pGEM-T* Easy vector

An Adenine (A) residue was attached to the 3' PCR products incubating at 72°C, along with TAQ DNA polymerase and the dNTPs for 30 min. The product size was then run on gel to check the fragment size. This fragment was then purified as described in Section 2.8. For the ligation process, the a-tailed PCR product was incubated overnight at 4°C with the *pGEM-T* Easy vector (Catalogue No.: A1360), following manufacturer's guidelines (LigaFast™ Rapid DNA Ligation System, Cat. No. - M8221) in a 10 µl final volume. The electroporation process was carried out using 20 µl of electrocompetent *E. coli* (Cat. No. EC100, Cambio Ltd.bacteria) cells and a 2 µl vector/ligation product.

The mixture was transferred to a 1mm gap-size cuvette (VWR Electroporation cuvettes Cat. No. 732-1135/ catalogue no. ECN:732-1135, 20 - 90 µl). The electroporator (Bio-rad, The Gene Pulser Xcell™ system in BTG lab) was set at 1.18v for bacteria according to the manufacturer's instructions. After removing the cuvette from the electroporator, 1ml of SOC medium (Cat. No. - S1797 Sigma SOC Medium) was immediately added to recover the bacteria from shock. This mixture was then transferred to a shaking incubator at 220rpm for an hour at 37°C. A total of four aliquots, 100µl, 150µl, 200µl and a final, concentrated one, were plated on LB (Cat. No. L3522 - Sigma LB Broth (Miller) plates by spreading the solution using a plate spreader. The plates were left in an incubator overnight at 37°C. The

next day selected colonies were screened using gene-specific primers to check the GOI insertion in the bacteria.

5.2.1.5 LB plate preparation

A mixture of suitable amounts of LB powder (Cat. No. L3522 - Sigma LB Broth (Miller) Powder microbial growth medium), agar (Agar No. 4 Plant Tissue Culture Grade Product Code: MC029, Lab M™) and double distilled water was weighed and autoclaved with a loosened cap. When the liquid's temperature was hand hot, suitable antibiotics were added in a sterile environment. LB plates containing 100µg/ml ampicillin (Cat. No. - A9518-5G Ampicillin sodium salt), 0.1 M IPTG (isopropyl/-Dthiogalactoside) (Cat. No. - I6758 Sigma-Aldrich IPTG; Cat. No. - BIO-37036 IPTG) and 20mg/ml X-Gal (Cat. No. - B4252 Sigma-Aldrich 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside) were used for growing *E. coli* electroporated bacteria. Around 25µl of this liquid was poured into petri dishes and left to dry for 20-30 minutes. The labelled and parafilm petri dishes were stored at 4°C.

5.2.1.6 Colony screening

After overnight incubation on LB and antibiotic plates, white colonies were selected and picked using pipette tips. The tip was rolled on fresh LB plate to produce more colonies, and the same tip was dipped in 50µl ddH₂O which was then used to amplify GOI using suitable primers through PCR. Primers, whether insert-specific (*EOMT*, internal primers), vector-specific (M13) or sometimes both, were used to screen the colonies. The thermocycler programme for M13 primers was an initial denaturation of two minutes at 94°C, 35 cycles of denaturation (94°C for 30 seconds), annealing (55°C for 30 seconds) and extension (72°C for 45 seconds), with a final extension at 72°C for two minutes. 5µl of amplification products were analysed on a 2 per cent agarose gel.

The thermocycler programme conditions for gene specific primers *EOMT*-F-1, *CVOMT*-F-1 and *OMT*-R-1074 were 94°C for two minutes for initial denaturation, followed by 94°C for 30 seconds, 52.6°C for 30 seconds, 72°C for 45 seconds for 35 cycles, and a final extension step set at 72°C for two minutes. Other gene-specific primers including *EOMT* and Internal primer pair's PCR conditions can be found in Chapter 4 M&M section.

5.2.1.7 Mini prep Plasmid DNA

A single colony was used to incubate in a 5ml LB medium containing suitable antibiotics and was kept overnight at 37°C on a shaker (250rpm). Next day this culture was centrifuged at 3000-4000 rpm for 10 mins. The plasmid DNA sample was collected in 50µl of elution buffer following the manufacturer's guidelines (QIAprep® Miniprep purification of high-copy plasmids, 2015).

5.2.1.8 Sequencing

After identifying positive colonies through PCR (with gene-specific and vector-specific primers (the PCR condition can be found in Section 5.2.1.6), relevant plasmid samples were sent for sequencing. Sequencing results could confirm the presence of GOI, correct orientation and junctions between the GOI and the plasmid.

5.2.2 Preparing PCR products using G-blocks

G-Blocks were also designed to be used in the transformation process. Two restriction sites (*EcoRI* and *NheI*) along with poly A tails were added to prepare four G-blocks. In the first instance, *pGEM-T* easy vector was used, where two REs were suitable: *NotI* (Promega, Catalog No.: R6431) and *EcoRI* (Promega, Catalog No.: R6011). Later, *pGEM* was replaced with the *pET* expression vector where *NotI* was replaced with *NheI* (Promega, Catalog No.: R6501). When the G-blocks were designed, all these restriction sites were therefore placed together as shown in Figure 5-4 and Figure 5-5) by designing a specific primer pair to make the fragment compatible with suitable vectors.

Poly A tail +Not1 + GOI + Ecor1+ Poly A tail

Figure 5-4: G-block design with relevant restriction enzymes and poly A tail (suitable for pGEM)

Poly A tail +Nhe1 + GOI + Ecor1+ Poly A tail

Figure 5-5: G-block design with modified restriction enzymes and poly A tail (suitable for pET)

A total of four G-block (gBlocks® Gene Fragments, IDT) were designed:

1. *O. basilicum* *EOMT* (from published sequence)
2. *O. basilicum* *CVOMT* (from published sequence)
3. *O. tenuiflorum* *EOMT* (from experiment sequence)
4. *O. gratissimum* *CVOMT* (from experiment sequence)

Full G-block sequence details can be found in Appendix.

O. basilicum *EOMT* (AF435008) and *O. basilicum* *CVOMT* (AF435007) sequences were obtained from the NCBI database.

As discussed in Chapter 4, a nucleotide alignment was prepared containing different *Ocimum* sequences. The remaining two G-blocks, *O. tenuiflorum* and *O. gratissimum* were designed using these nucleotide alignments. The full-length DNA sequences of *O. tenuiflorum* and *O. gratissimum* are available in Appendix.

5.2.3 *pET* vector cloning

pET-28a (+) vector (Catalog No. – 69864-3, Merck) strain was used to design protein expression assays (Figure 5-6).

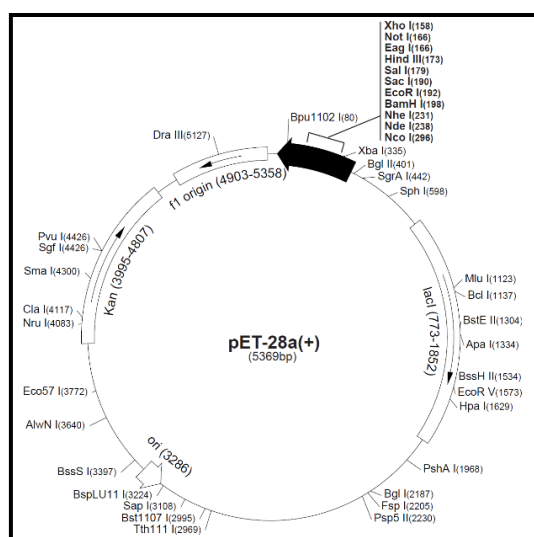


Figure 5-6: *pET-28a(+)* Vector Map

5.2.3.1 Preparing G-blocks for PCR using *pET* primers

G-blocks were diluted as follows to create 1/10, 1/100, 1/500 and 1/1000 dilutions, and 1/500 G-blocks were eventually used for PCR with newly designed *pET* primers (Table 5-3). The thermocycler programme was set at 94°C for two minutes for initial denaturation, 94°C for 30 seconds, followed by five cycles of denaturation at 48°C, then 30 cycles of denaturation at 58°C, 72°C for 45 seconds, with the final extension step set at 72°C for two minutes. All PCR products were visualised using 2% w/v agarose gel.

5.2.3.2 PCR purification

After replacing RE sites using *pET* primers, PCR reactions were purified using a QIAquick PCR Purification Kit (Qiagen, Cat No./ID: 28104).

5.2.3.3 Restriction digestion of Gene of Interest (GOI) and vector

After adding *pET* cloning sites through PCR, all four G-blocks were digested with *NheI* and *EcoRI*. 20µl of reaction volume was prepared by using 2µl 10X NEBuffer, 0.2µl BSA, 0.5µl *NheI*, 0.5µl *EcoRI* and 1µg PCR product. The digested reaction samples were incubated at 37°C for an hour and all products were analysed by agarose gel electrophoresis. Afterwards, the right size bands were gel-purified using the QIAquick Gel Extraction Kit (Cat No./ID: 28704).

5.2.3.4 Preparing the ligation product

The *pGEM* ligation kit (Promega, Catalog number: A1360) was used for *pET* vector ligation where the *pGEM* vector was used as a positive vector. Digested G-blocks' PCR products were used as a template for the ligation process. The latter were prepared using the manufacturer's instructions, with a total volume of 20µl.

Apart from using 50mg/mL of Kanamycin (Cat. No. - K0254 Sigma Kanamycin solution from *Streptomyces kanamyceticus*) to prepare LB plates, all transformation steps were carried out as per *pGEM* electroporation (see Section 5.2.1).

5.2.3.5 *pET* colony screening

T7 promoter and T7 reverse primer pairs were used to detect if the full-length sequence had been inserted into the *pET* vector. The gene-specific primers *EOMT*, *EOMT* (1-1074) and Internal primers, alone or in combination with T7 primers, were also used to screen colonies.

There were two thermocycler programmes designed for PCR reaction with specially designed *pET* vector primers. The thermocycler programme was first set with two annealing temperatures: 48°C for an initial five cycles, then 58°C for 30 cycles, while for the second PCR programme only one annealing temperature of 58°C was set. The complete PCR programme is described in Table 5-2.

Table 5-2: Different thermocycler programmes designed for pET primers

pET primers PCR program	Thermocycler conditions
Thermocycler programme 1	An initial denaturation of 2 minutes at 94°C, 5 cycles of denaturation (94°C for 30 seconds), annealing (48°C for 30 seconds) and extension (72°C for 45 seconds); followed by 30 cycles of denaturation (94°C for 30 seconds), annealing (58°C for 30 seconds) and extension (72°C for 45 seconds); final extension of 72°C for 2 minutes.
Thermocycler programme 2	An initial denaturation of 2 minutes at 94°C, 35 cycles of denaturation (94°C for 30 seconds), annealing (58°C for 30 seconds) and extension (72°C for 45 seconds); final extension of 72°C for 2 minutes.

Finally, thermocycler programme 1 was selected for *pET* primers.

T7 primers details can be found in table 5-3.

Table 5-3: Vector-specific and gene-specific primer list

Primers used for pET colony screening	Sequence	Thermocycler programme
<i>pET_EOMT-F</i>	AAAAAAGCTAGCATGGCATTGCAAAAAGTA	An initial denaturation of 2 minutes at 94°C, 5 cycles of denaturation (94°C for 30 seconds), annealing (48°C for 30 seconds) and extension (72°C for 45 seconds); followed by 30 cycles of denaturation (94°C for 30 seconds), annealing (58°C for 30 seconds) and extension (72°C for 45 seconds); final extension of 72°C for 2 minutes.
<i>pET_CVOMT-F</i>	AAAAAAGCTAGCATGGCATTGCAAAATATG	
<i>pET_OMT-R</i>	TTTTTTGAATTCTTAAGGATAAGCCTCTAT	
T7_promoter	TAATACGACTCACTATAG	An initial denaturation of 2 minutes at 94°C, 35 cycles of denaturation (94°C for 30 seconds), annealing (53°C for 30 seconds) and extension (72°C for 45 seconds); final extension of 72°C for 2 minutes.
T7_terminator	GCTAGTTATTGCTCAGCGG	

Amplification products (5µl) were analysed on a 2% w/v agarose gel. Positive colonies were grown overnight in 5ml of liquid LB for preparing plasmid DNA samples. The plasmid DNA were sent to sequence to check the correct orientation of gene ligate into the vector.

5.2.4 Sequence analysis

The CLC bio main workbench (CLC bio, Massachusetts USA) was used to study sequences of both *pGEM* and *pET* colony screening. Additionally, the NCBI BLAST search tool was used to identify the relevant gene presence in the colony. Gene orientation was also checked using reference/published *EOMT* and *CVOMT* sequences.

5.2.5 IPTG Assay

5ml of liquid culture containing LB and 50mg/mL Kanamycin was inoculated with selected colonies streaked onto fresh LB plates with 50mg/mL of Kanamycin). This culture was incubated overnight at 37°C in a shaker set at 220rpm. The next day 200µl of this culture was used to inoculate 20mL of fresh LB culture. When a cultural density of 0.4 to 0.5 OD600 was reached using a spectrophotometer, protein expression was induced by 0.3mM IPTG. This culture was kept overnight on a shaker at 220rpm at 20°C. The next day, the cell pellet was collected by centrifugation at 3,500 rpm for 10 minutes at 4°C and stored at -20°C. The lysis buffer was prepared, containing 10% Glycerol, 50mM bis-Tris, 1mM EDTA, 14mM 2-mercaptoethanol, 10mM NaCl and 1µg/mL lysozyme in 500 µl solution. The lysed cells were then incubated for 20 minutes on ice and liquid nitrogen and thawed a couple of times. The centrifugation was carried out at 13,000 rpm for 10 minutes at 4°C to remove the cell debris. Further purification and enzymatic assays were designed using this crude bacterial lysate.

5.3 Result

5.3.1 Primers covering initial and end parts of the OMT gene region

The majority of young *Ocimum* leaves were collected for the RNA extraction procedure as described in Section 2.4). cDNA containing partial sequences of the *OMT* gene was synthesised from this extracted RNA by using PCR specific primers. NCBI BLAST search tool was used to identify similar sequences. As discussed earlier, two published *OMT* sequences were obtained from the database: *O. basilicum EOMT* (AF435008) and *O. basilicum CVOMT* (AF435007). These sequences contain the ATG start codon and the STOP

codon. After aligning these reference sequences with other experimental sequences, differences in the nucleotide sequence were noticed in the beginning part of the sequences. As shown in figure 5-7, a difference of three bases can be seen between the initial *EOMT* and *CVOMT* sequences. This is why two forward primers were designed (see table 5-4), each of which can anneal to the first parts of these sequences.

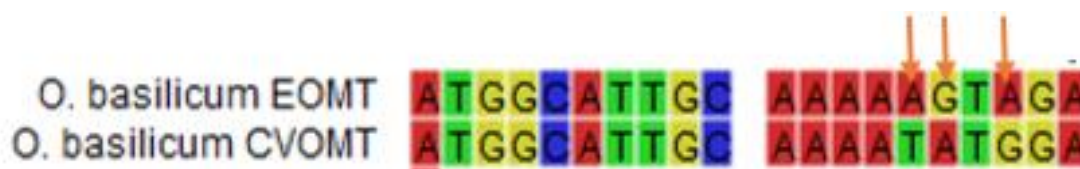


Figure 5-7: The base pair differences between the first parts of the published *O. basilicum* *EOMT* and *CVOMT* sequences. Arrows show the difference in the nucleotides sequence between the two *OMTs*

Yet no differences were noted at the 5' end. Therefore, only one reverse primer, which can anneal to both reference sequences, was therefore designed. The reverse primer sequence is detailed in table 5-4.

Table 5-4: *EOMT* and *CVOMT* forward primers (called *EOMT*-F-1 and *CVOMT*-F-1), *OMT* reverse primer (called *OMT*-R-1074)

<i>EOMT</i> -F-1	5'-ATGGCATTGCAAAAAGTAGA-3'
<i>CVOMT</i> -F-1	5'-ATGGCATTGCAAAAATATGGA-3'
<i>OMT</i> -R-1074	5'-TTAAGGATAAGCCTCTATGAG-3'

Using the two set of forward primers and the reverse primer, the full lengths (about 1,100bp) of both *CVOMT* and *EOMT* were amplified using genomic DNA from different *Ocimum* samples. Figure 5-8 shows that it was possible to obtain the full length of *EOMT* sequence from *O. gratissimum* (lane 2) and *O. basilicum* (lane 3) using the *EOMT*-F-1 and *OMT*-R primer pair, while the full length of *CVOMT* sequence from *O. kilimandscharicum* (lane 4), *O. basilicum* (lane 5) and *O. americanum* (lane 6) using the *CVOMT*-F-1 and *OMT*-R primer pair. The band was present for positive control (lane 1), and no bands were visible in the negative control (lane 7).

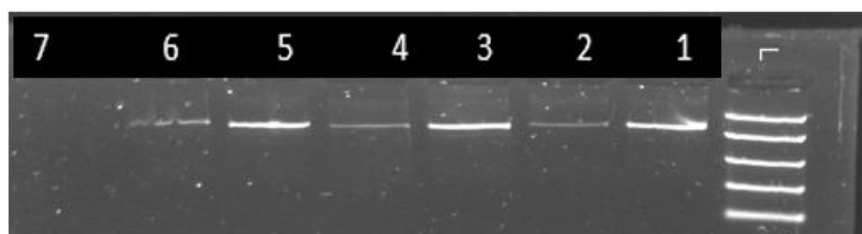


Figure 5-8: PCR analysis of different *Ocimum* genomic DNA samples using *EOMT*-F-1 and *CVOMT*-F-1 in combination with the *OMT*-R-1074 primer pair. L: DNA ladder; Lane 1: positive control; Lane 2: *O. gratissimum*; Lane 3: *O. basilicum* samples as template with *EOMT*-F-1 and *OMT*-R primers; Lane 4: *O. kilimandscharicum*; Lane 5: *O. basilicum*; Lane 6: *O. americanum* DNA samples as template with *CVOMT*-F-1 and *OMT*-R primer pairs; Lane 7: negative control

5.3.2 *pGEM* vector cloning

5.3.2.1 RNA extraction from *O. tenuiflorum* species

Two RNA samples were isolated from an *O. tenuiflorum* plant. The 260/280 ratio checked with the nanodrop was 0.176 and 0.116 respectively. The quality of the RNA was also checked by running 5ul of RNA with loading dye (1:5 ratio) on 1% w/v agarose gel. Two *O. tenuiflorum* RNA samples can be seen in Lanes 1 and 2 (Figure 5-9). Figure 5-9 (b) shows the band of genomic DNA (Lane 1), RNA (Lane 2) and cDNA product (Lane 3). No bands were visible in the negative control (Lane 4). The expected product sizes are present in both gel pictures.

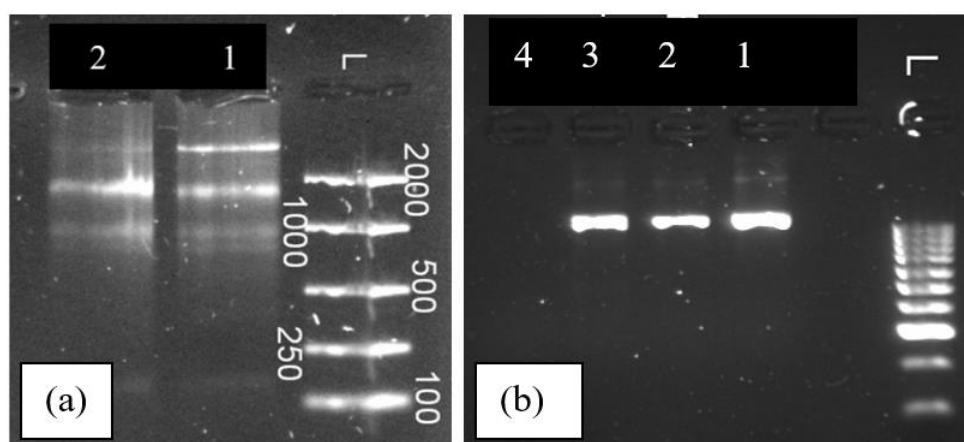


Figure 5-9: RNA extraction samples. (a) L: DNA ladder; Lanes 1 & 2: *O. tenuiflorum* RNA samples. (b) L: DNA ladder; Lane 1: *O. tenuiflorum* genomic DNA sample; Lane 2: *O. tenuiflorum* RNA samples; Lane 3: c-DNA sample; Lane 4: negative control

5.3.2.1.1 Turbo DNA clean-up and cDNA synthesis

As genomic DNA traces were found after the RNA extraction procedure, a further clean-up process, a turbo DNA clean-up, was carried out. In figure 5-10, the band can be seen for the RNA (Lane 1) and cDNA (Lane 3) samples, while no bands are visible in the negative control (Lane 4). The expected product sizes are present in the gel picture.

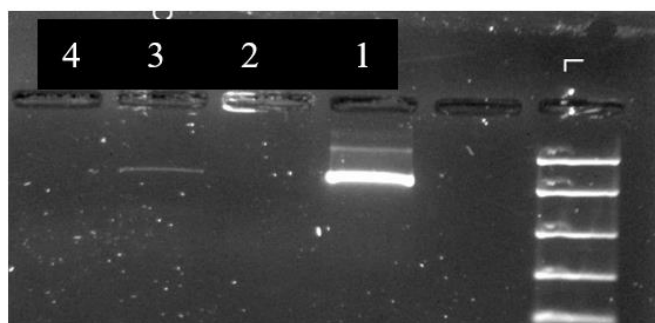


Figure 5-10: Comparing Genomic DNA, Turbo DNA clean up and cDNA samples. L: DNA ladder; (+), Lane 1: positive control; Lane 2: turbo DNA sample; Lane 3: cDNA sample and Lane 4 negative control

5.3.2.2 Blue-white screening for *pGEMT* Easy vector

Around 1,100bp gel extracted products were ligated to *pGEMT* Easy vector to transform electrocompetent *E. coli* cells. Ligation products were confirmed through PCR reaction as shown in figure 5-11, where the positive control (Lane 1) and the ligation product (Lane 3) shows bands, while no bands are present in the negative control (Lane 2). The expected product sizes are present in the gel picture.

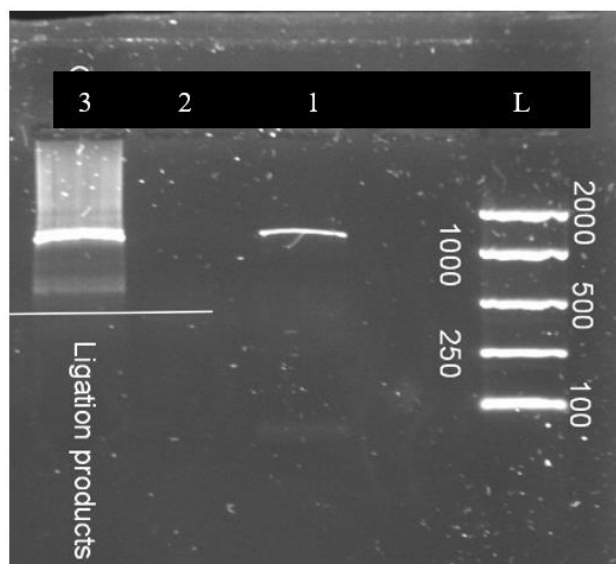


Figure 5-11: PCR reaction confirmation of ligation products. L: DNA ladder; Lane 1: positive control; Lane 2: negative control; Lane 3: GOI.

5.3.2.3 *pGEM* Colony screening

Vector-specific M13 primers were used to check gene orientation, as well as checking for the GOI presence in specific colonies. As shown in figure 5-12 (a), there were bands present in 14 white colonies which were PCR with M13 primers (Lanes 4 to 17), along with one blue colony (Lane 2) and the standard control (Lane 3). No bands were present in negative control (Lane 1). Two of these 14 colonies, the fourth and fourteenth, were selected for sequencing. In figure 5-12 (b), the bands are present in the blue, (Lane 1) fourth (Lane 2) and fourteenth colonies (Lane 3), while no bands were present in the negative control (Lane 4). The expected product sizes are present in both gel pictures.

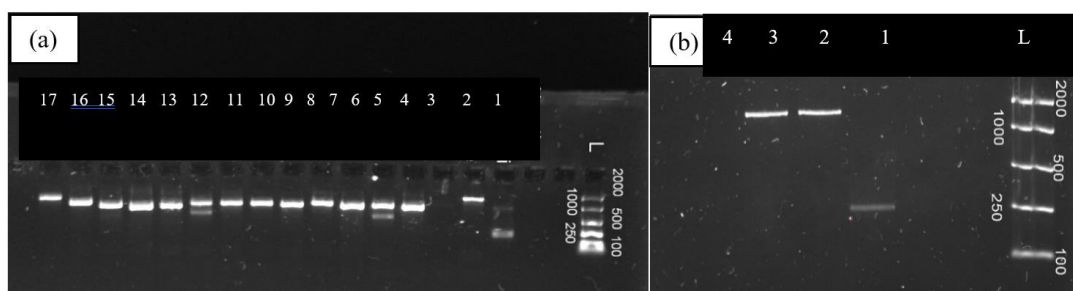


Figure 5-12: pGEM colony screening using M13 primers (a) Total of 14 colonies screened. L: DNA ladder; Lane 1: negative control; Lane 2: Blue colony control; Lane 3: Standard white colony control: Lanes 4-17: transformed colonies. (b) PCR confirmation of selected colonies. L: DNA ladder; Lane 1: blue colony; Lane 2: fourth colony; Lane 3: fourteenth colony; Lane 4: negative control

5.3.2.3.1 Plasmid prep and sequencing of selected pGEM colonies

The miniprep procedure was carried out for the fourth and fourteenth colonies to extract plasmid DNA samples from them. These samples were then sent for sequencing with M13 primers. As shown in figure 5-13, the GOI's presence has been confirmed using the BLAST tool.

<input type="checkbox"/>	Ocimum tenuiflorum biotype 2 eugenol O-methyltransferase mRNA, complete cds	1604	1604	74%	0.0	97%	EU622049.1
<input type="checkbox"/>	Ocimum basilicum eugenol O-methyltransferase mRNA, complete cds	1288	1288	82%	0.0	88%	AF435008.1
<input type="checkbox"/>	Ocimum basilicum CVOMT mRNA for chavicol O-methyltransferase, complete cds	1275	1275	80%	0.0	89%	AB530137.1
<input type="checkbox"/>	Ocimum basilicum chavicol O-methyltransferase mRNA, complete cds	1264	1264	80%	0.0	89%	HM990155.1
<input type="checkbox"/>	Ocimum tenuiflorum biotype 6 eugenol O-methyltransferase gene, complete cds	1242	1606	75%	0.0	98%	EU622046.1
<input type="checkbox"/>	Ocimum tenuiflorum biotype 8 eugenol O-methyltransferase gene, complete cds	1236	1606	73%	0.0	98%	EU622048.1
<input type="checkbox"/>	Ocimum tenuiflorum biotype 5 eugenol O-methyltransferase gene, complete cds	1236	1618	75%	0.0	98%	EU622045.1
<input type="checkbox"/>	Ocimum tenuiflorum biotype 1 eugenol O-methyltransferase gene, complete cds	1236	1606	73%	0.0	98%	EU622042.1
<input type="checkbox"/>	Ocimum basilicum chavicol O-methyltransferase mRNA, complete cds	1236	1236	80%	0.0	88%	AF435007.1
<input type="checkbox"/>	Ocimum tenuiflorum biotype 2 eugenol O-methyltransferase gene, complete cds	1230	1595	72%	0.0	98%	EU622043.1
<input type="checkbox"/>	Ocimum tenuiflorum biotype 3 eugenol O-methyltransferase gene, complete cds	1225	1590	73%	0.0	98%	EU622044.1
<input type="checkbox"/>	Ocimum tenuiflorum biotype 7 eugenol O-methyltransferase gene, complete cds	1218	1590	71%	0.0	97%	EU622047.1

Figure 5-13: Screenshot of the BLAST search to identify the GOI's presence in the colony sequencing results. The BLAST hits reflect the percentage of identity. The table contains all hits for the query together with the associated BLAST statistics.

Moreover, gene orientation has also been confirmed by identifying the initial sequence in the results. This sequence has been highlighted in figure 5-14.

Sequences of positive colonies revealed the presence of *O- methyltransferase* genes. A sequence of 1,074bp along with a few extra bp products containing poly A tails were generated by PCR during both *pGEM* vector transformations.

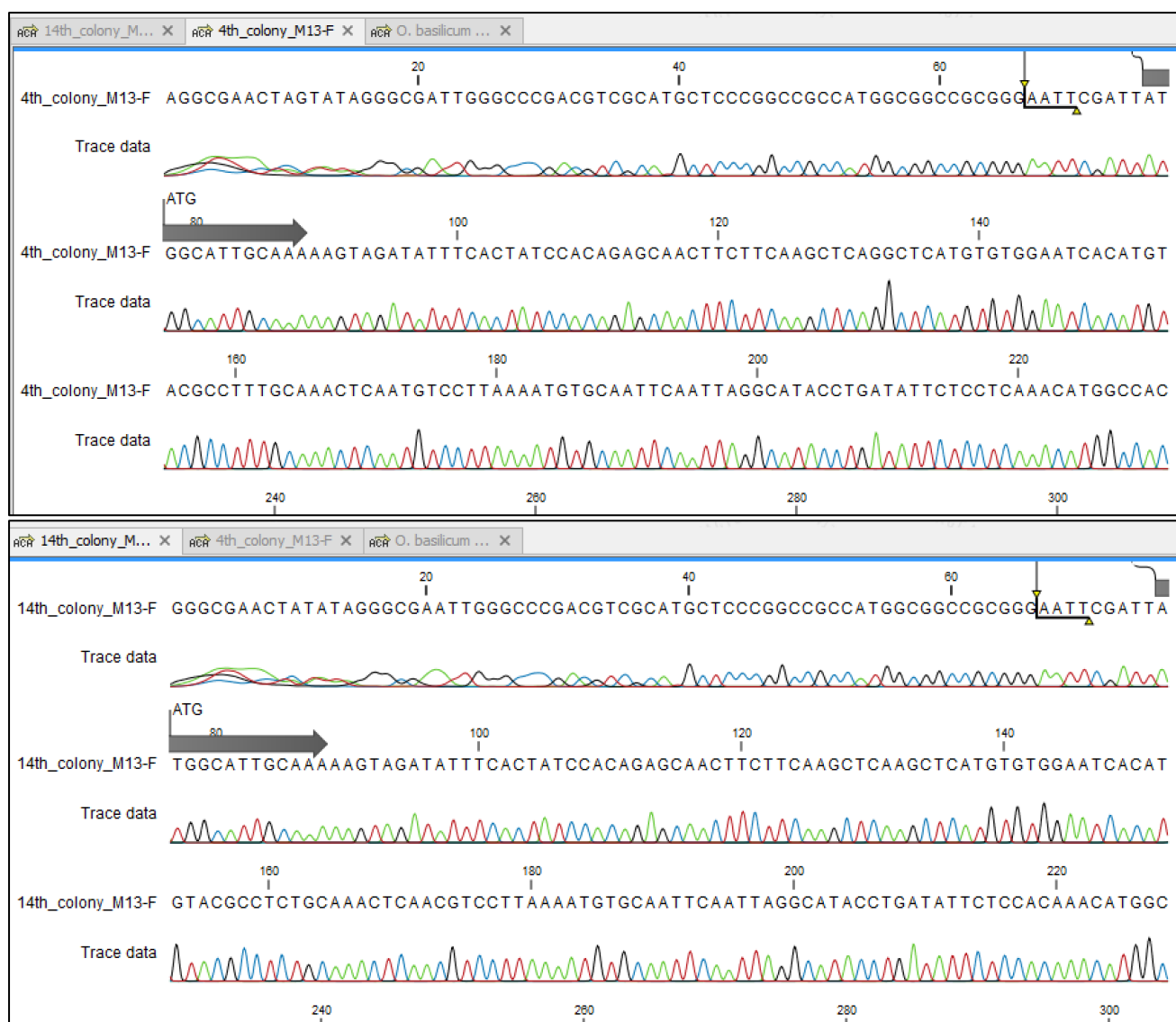


Figure 5-14: Chromatogram of annotated regions showing the correct GOI orientation in the fourth and fourteenth *pGEM* colonies

5.3.3 G-blocks preparation for *pET* vector cloning

The following steps were performed to add suitable RE sites before transforming the *pET* vector using four G-blocks.

1. Diluting the G-blocks
2. Replacing the *NotI* RE site with *NheI* using *pET* primers
3. Purifying the PCR samples
4. Digesting the purified PCR samples with *NheI* and *EcoRI* RE
5. Gel extraction of the correct size band or PCR purification

5.3.3.1 Diluting G-blocks

The G-blocks were suspended by following the manufacturer's guidelines before dilution. As G-block aliquots are quite concentrated, a series of dilutions was prepared before using them for PCR as a template. As shown in figure 5-15, 1/100, 1/500, 1/1000 and 1/5000 dilutions were tested while using them with *pET* primers.

5.3.3.2 Replacing the *NotI* RE site with *NheI* using *pET* primers

The reason for designing this PCR was to replace the *NotI* Restriction site (which is present in G-blocks) with *NheI* (which is present in *pET* vector MCS). The *NotI* restriction site is not present in the *pET* vector MCS.

5µl of amplification products were analysed on 2% w/v agarose gel. As shown in figure 5-15, the first thermocycler program gave better results than the second. Figure 5-15 (a) represents the PCR results for thermocycler program 1, where four bands can be seen in Lanes 2, 3, 4 and 5 representing *O. basilicum_EOMT*, *O. basilicum_CVOMT*, *O. tenuiflorum_EOMT* and *O. graticimum_EOMT* G-blocks respectively. No bands were present in the negative control (Lane 6). Figure 5-15 (b) represents the PCR results for thermocycler program 2, where faint bands were present in Lanes 2, 3, 4 and 5 representing *O. basilicum_EOMT*, *O. basilicum_CVOMT*, *O. tenuiflorum_EOMT* and *O. graticimum_EOMT* G-blocks respectively. The expected product sizes are present in both gel pictures.

Both sets of *pET_EOMT* and *pET_CVOMT* primers were tested with all four G-blocks (*O. basilicum_EOMT*, *O. basilicum_CVOMT*, *O. tenuiflorum_EOMT* and *O. graticimum_EOMT*) and as expected, the *pET_EOMT* primers worked with all three *EOMT* G-blocks (*O. basilicum_EOMT*, *O. tenuiflorum_EOMT* and *O. graticimum_EOMT*), while the *pET_CVOMT* primers worked with one *CVOMT* G-block (*O. basilicum_CVOMT*).

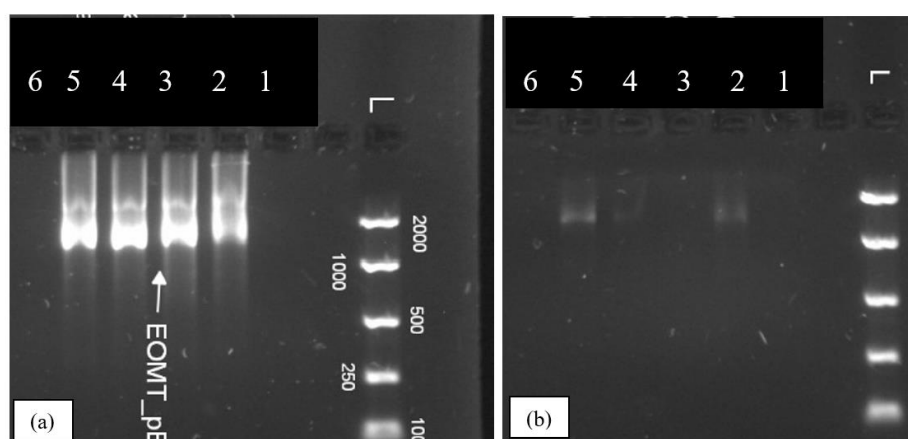


Figure 5-15: Two thermocycler programmes designed for pET *EOMT* primers using G-blocks (1/500 and 1/100 dilutions) as template, as per Table 5-2. (a) Thermocycler programme 1. L: DNA ladder; Lane 1: positive control; Lane 2: *O. tenuiflorum* (1/500) G-block; Lane 3: *O. tenuiflorum* (1/100) G-block; Lane 4: *O. gratissimum* (1/500) G-block; Lane 5: *O. gratissimum* (1/100) G-block; Lane 6: negative control. (b) Thermocycler programme 2. L: DNA ladder; Lane 1: positive control; Lane 2: *O. tenuiflorum* (1/500) G-block; Lane 3: *O. tenuiflorum* (1/100) G-block; Lane 4: *O. gratissimum* (1/500) G-block; Lane 5: *O. gratissimum* (1/100) G-block; Lane 6: negative control.

These *pET* primers were later cross-checked using G-blocks as templates. In figure 5-16, bands were visible in Lanes 2, 3, 4 and 5 representing *O. tenuiflorum* (1/500), *O. tenuiflorum* (1/100), *O. gratissimum* (1/500) and *O. gratissimum* (1/100) G-blocks. No bands were visible in the negative control (Lane 6). As expected, no bands were visible using the same templates (*O. basilicum*, *O. tenuiflorum* and *O. gratissimum*) with pET_*CVOMT* in Lanes 7 to 12.

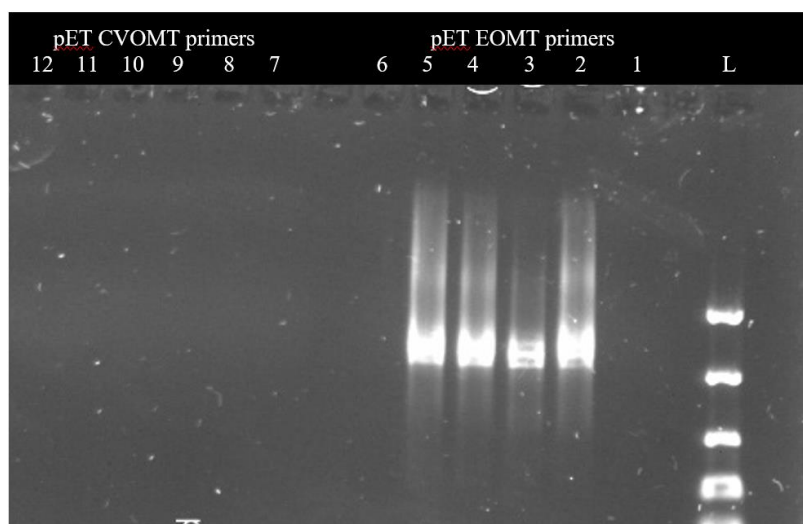


Figure 5-16: *pET_EOMT* and *pET_CVOMT* primers were cross checked with each G-block template. The PCR product of the pET_*EOMT* primer was: L: DNA ladder; Lane 1: positive control; Lane 2: *O. tenuiflorum* (1/500) G-block; Lane 3: *O. tenuiflorum* (1/100) G-block; Lane 4: *O. gratissimum* (1/500) G-block; Lane 5: *O. gratissimum* (1/100) G-block; Lane 6: negative control. PCR product of pET_*CVOMT* primer; Lane 7: positive control; Lane 8: *O. tenuiflorum* (1/500) G-block; Lane 9: *O. tenuiflorum* (1/100) G-block; Lane 10: *O. gratissimum* (1/500) G-block; Lane 11: *O. gratissimum* (1/100) G-block; Lane 12: negative control.

5.3.3.3 Purifying PCR samples

After replacing RE sites using *pET* primers, PCR products were prepared for restriction digestion through a purification process. These samples were further checked by using *pET* primers. Figure 5-17 (a) shows that Lanes 2, 3 and 4 represent the bands for *EOMT* G-blocks. No bands were present in the negative control (Lane 5). In figure 5-17 (b), Lane 2 shows a *CVOMT* G-block, while no bands were present in the negative control (Lane 3). The expected product sizes are present in both gel pictures.

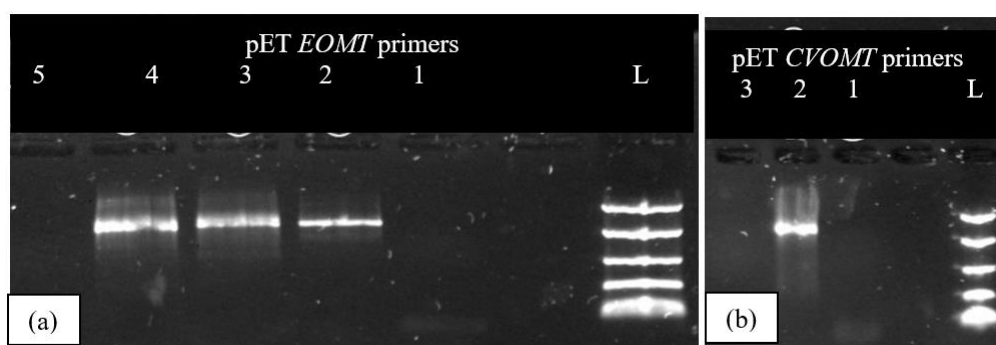


Figure 5-17: Purified PCR samples containing all four G-blocks. (a) PCR using pET *EOMT* primers. L: DNA ladder; Lane 1: positive control; Lane 2: *O. basilicum* (*EOMT*) G-block; Lane 3: *O. tenuiflorum* G-block; Lane 4: *O. gratissimum* G-block; Lane 5: negative control. (b) PCR using pET *CVOMT* primers. L: DNA ladder; Lane 1: positive control; Lane 2: *O. basilicum* (*CVOMT*); Lane 3: negative control

5.3.3.4 Digesting purified PCR samples with *NheI* and *EcoRI* RE

The purified PCR samples were digested using two different REs: *NheI* at 5' end and *EcoRI* at 3' end. The digested products can be seen in Figure 5-18 (Lanes 5, 4, 3 and 2), while the digested *pET* vector can be seen in Lane 1. The expected product sizes were present in gel picture. These four G-block RE bands were gel-purified for further processing.

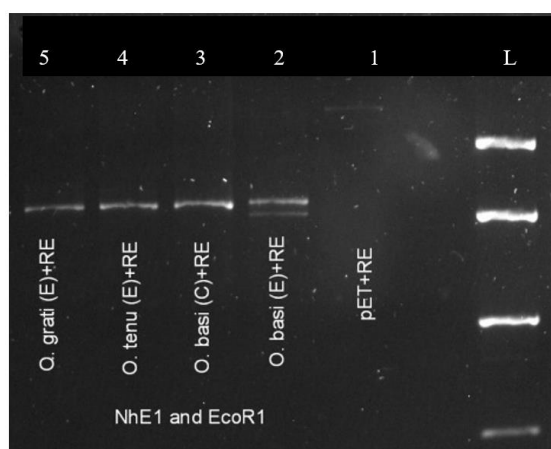


Figure 5-18: Restriction digest product of all G-blocks and *pET* vector using *NheI* and *EcoRI* RE. L: DNA ladder; Lane 1: *pET* vector; Lane 2: *O. basilicum* (*EOMT*) G-block; Lane 3: *O. basilicum* (*CVOMT*) G-block; Lane 4: *O. tenuiflorum* G-block; Lane 5: *O. gratissimum* G-block.

5.3.4 *pET* colony screening

Initially, published *EOMT* primers were used to identify common *OMT* genes from different *Ocimum* species (see the results). Later, a pair of universal T7 primers were used in combination with *pET* and *OMT* primers.

5.3.4.1 PCR using gene-specific and vector-specific primers

After growing colonies overnight, recombinant colonies were screened using PCR for quick identification. Gene-specific (*EOMT*) primers were used to screen recombinant plasmids (Figure 5-19). Figure 5-19 (a) shows the bands for positive control. *O. basilicum* colonies 1 to 8 had shown the bands in Lanes 1 to 8, three *O. tenuiflorum* colonies in Lanes 9, 10 and 11 and three *O. gratissimum* colonies in Lanes 12, 13 and 14. No bands were present in Lane 15, implying negative control. Figure 5-19 (b) also shows the bands for positive control. *O. basilicum* (*CVOMT*) colonies 4, 5, 6, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 23 and 24 had shown their bands in the respective lanes. No bands were present in the negative control (Lane 4). The expected product sizes were present in both gel pictures.

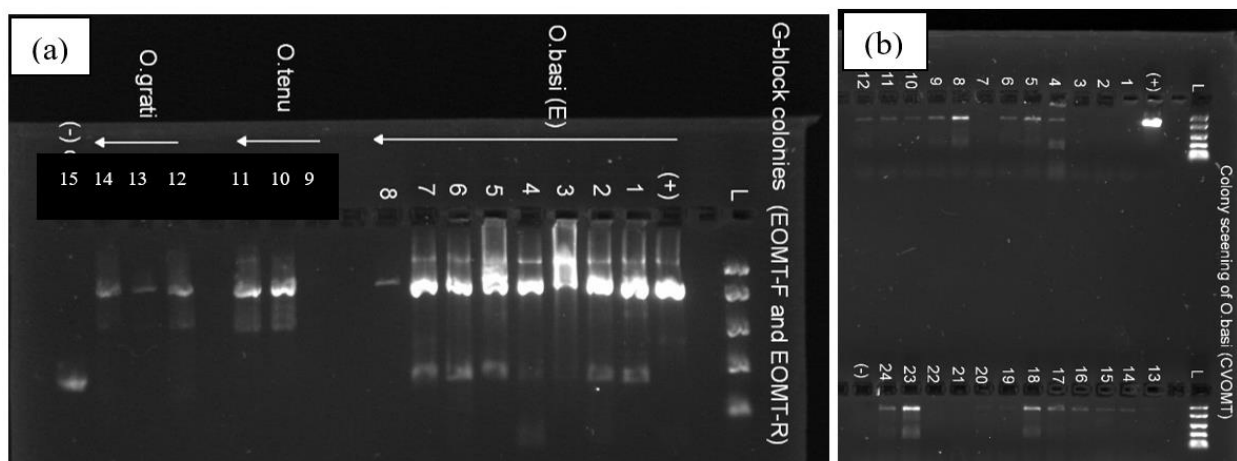


Figure 5-19: pET vector PCR colony screening using *EOMT* (gene-specific) primers. (a) Colony screening PCR results of *O. basilicum* (*EOMT*), *O. tenuiflorum* and *O. gratissimum*. L: DNA ladder; (+): positive control; Lanes 1-8: *O. basilicum* (*EOMT*) colonies; Lanes 9-11: *O. tenuiflorum* colonies; Lanes 12-14: *O. gratissimum* colonies; Lane 15: negative control. (b) Colony screening PCR results of *O. basilicum* (*CVOMT*). L: DNA ladder; (+): positive control; Lanes 1-24: *O. basilicum* (*CVOMT*) colonies; (-): negative control

The colony screening led to one colony from each G-block and *pET* transformation being selected for further processing. As shown in figure 5-20, the second colony of *O. basilicum* (*EOMT*), the 23rd of *O. basilicum* (*CVOMT*), the third of *O. tenuiflorum* and the first of *O. gratissimum* were selected for mini prep plasmid preparation. These plasmids were PCR-

checked using *EOMT* primers before being sent for sequencing. The expected size (1,074bp) bands were present in all wells including the positive control (Lane 1), the second colony of *O. basilicum* (*EOMT*) (Lane 2), the 23rd of *O. basilicum* (*CVOMT*) (Lane 3), the third of *O. tenuiflorum* (Lane 4) and the first of *O. gratissimum* (Lane 5).

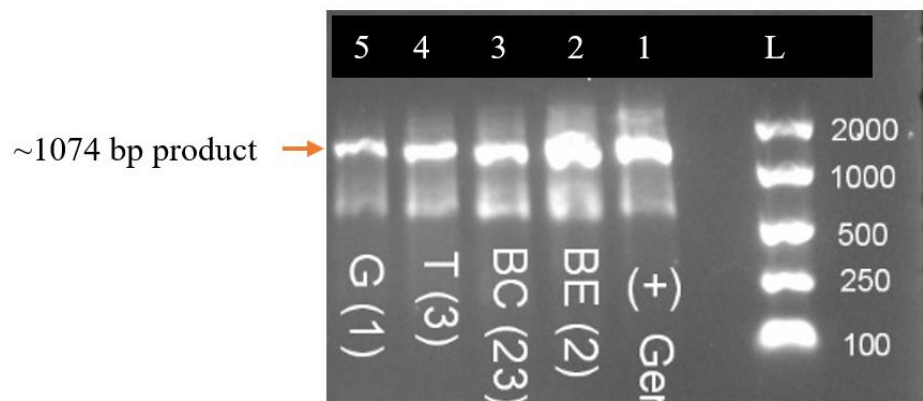


Figure 5-20: PCR test confirmation for selected pET colonies using *EOMT* primers. L: DNA ladder; Lane 1: positive control; Lane 2: 2nd colony of *O. basilicum* (*EOMT*); Lane 3: 23rd colony of *O. basilicum* (*CVOMT*) and 3rd colony of *O. tenuiflorum*; Lane 4: 1st colony of *O. gratissimum*; Lane 5: negative control

5.3.4.2 Sequence analysis of selected *pET* colonies

As shown in figure 5-21, 5-22, 5-23 and 5-24, the presence of the *Ocimum* sequence was confirmed using a GOI orientation check. Based on this finding, initial region showed a starting codon followed by some base pair differences (apart from *O. gratissimum*) because of poor quality sequencing. The regions are annotated and base pair differences are highlighted.

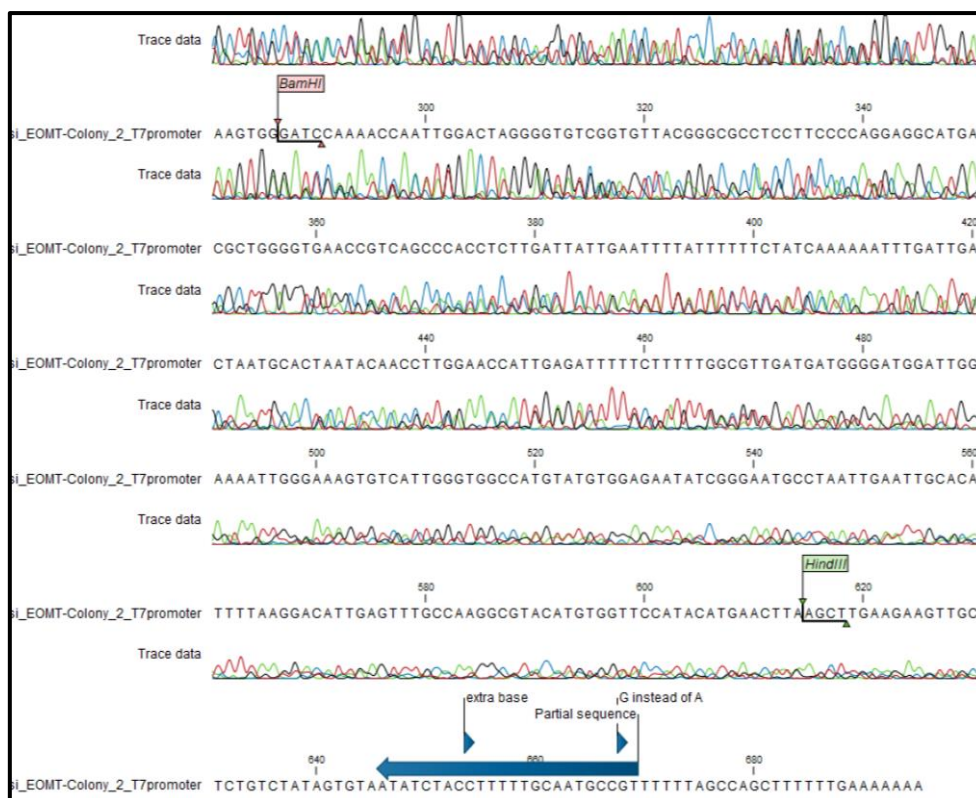


Figure 5-21: Chromatogram with annotated regions showing the correct GOI orientation (with blue arrow showing the starting codon) in the 2nd colony of the *O. basilicum* (EOMT) sequence analysis

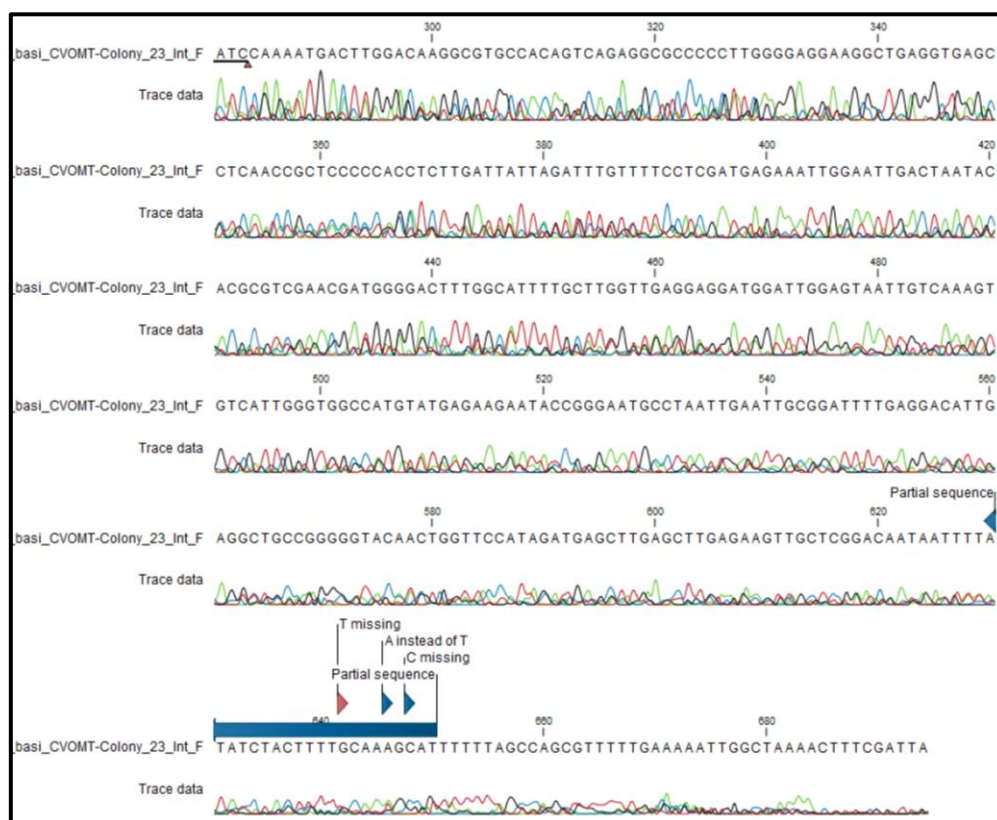


Figure 5-22: Chromatogram with annotated regions showing the correct GOI orientation (with blue arrow showing the starting codon) in the 23rd colony of the *O. basilicum* (CVOMT) sequence analysis

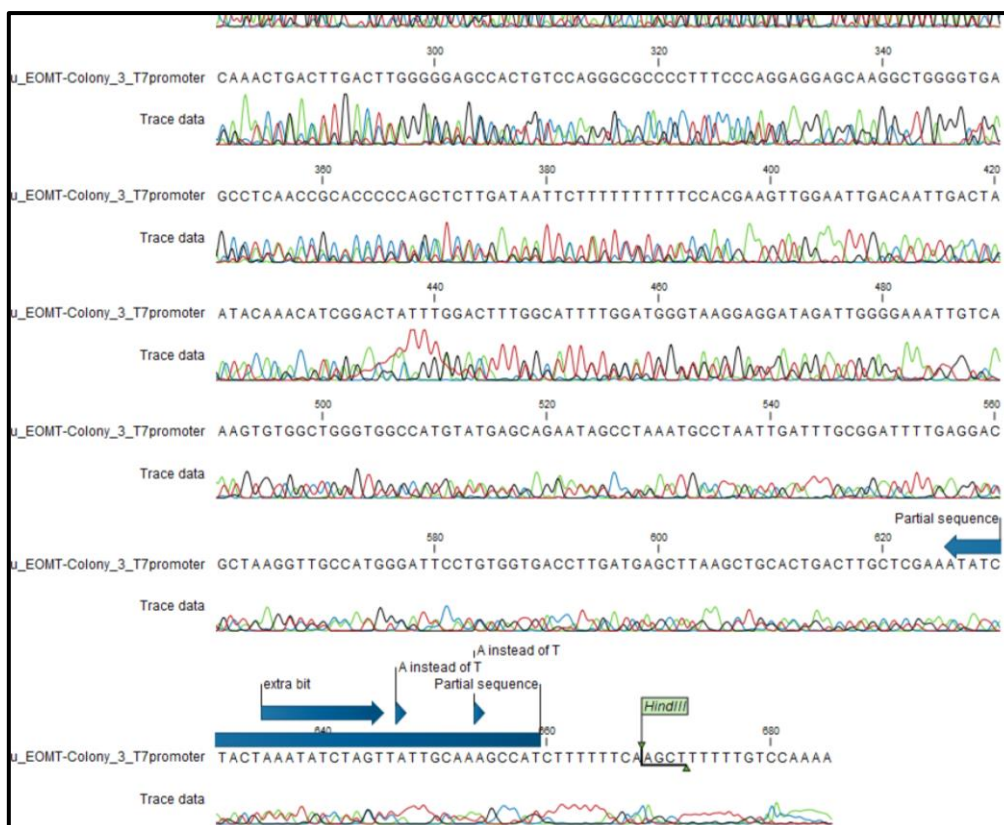


Figure 5-23: Chromatogram with annotated regions showing the correct GOI orientation (with blue arrow showing the starting codon) in the 3rd colony of the *O. tenuiflorum* sequence analysis

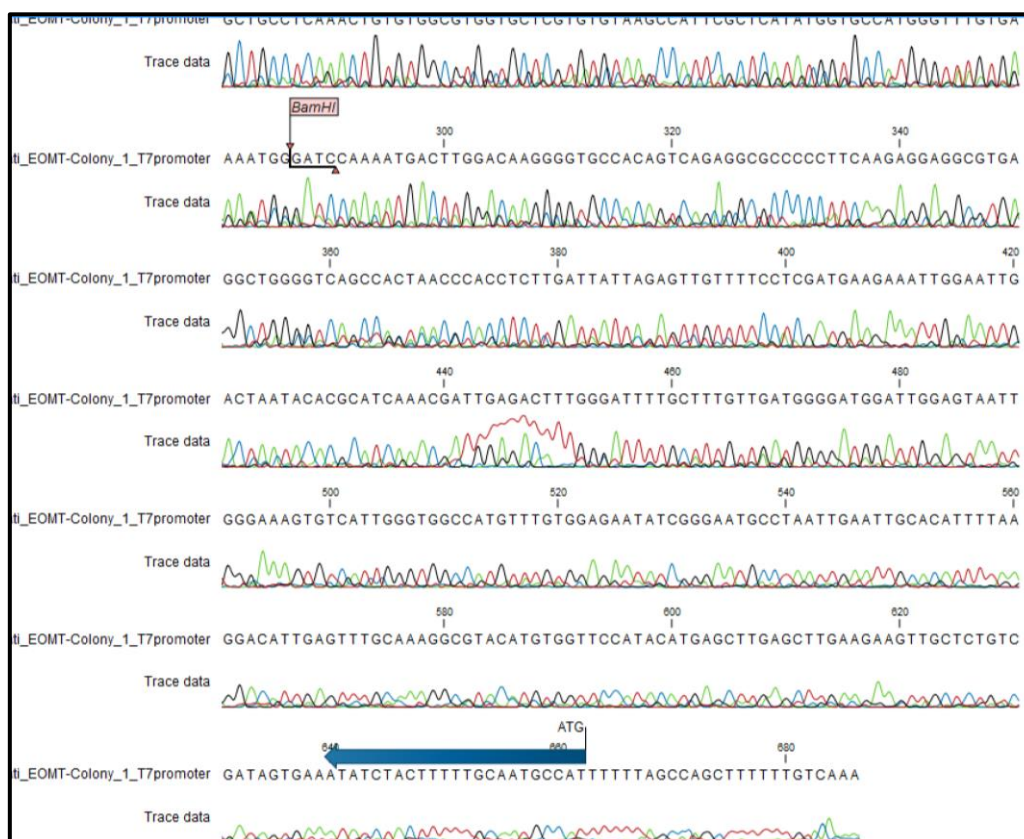


Figure 5-24: Chromatogram with annotated regions showing the correct GOI orientation (with blue arrow showing the starting codon) in the 1st colony of the *O. gratissimum* sequence analysis

5.3.5 SDS-PAGE

As mentioned in the introduction, the main aim of using the SDS-PAGE technique is to separate and characterize proteins. Four main protein samples were used: *O. basilicum* (EOMT), *O. basilicum* (CVOMT), *O. tenuiflorum* and *O. gratissimum*, along with the *pET* vector as a control. Previous research indicates that both EOMT and CVOMT subunits have a molecular mass of ~40kD (Ibrahim *et al.*, 1998; Wang, 1999). As no information is available whether these proteins are in solution or pellet form, both supernatant and pellets were examined on the gel along with the protein ladder to determine their size (Figure 5-25 and Figure 5-26). Two samples were prepared, one at low OD (around 0.3-0.4) (in Lane 1, 2, 3, 4 and 5) and the other at high OD (around 0.5-0.6) (in Lane 6, 7, 8, 9 and 10). The samples at 40kD are highlighted in figure 5-25 and 5-26.

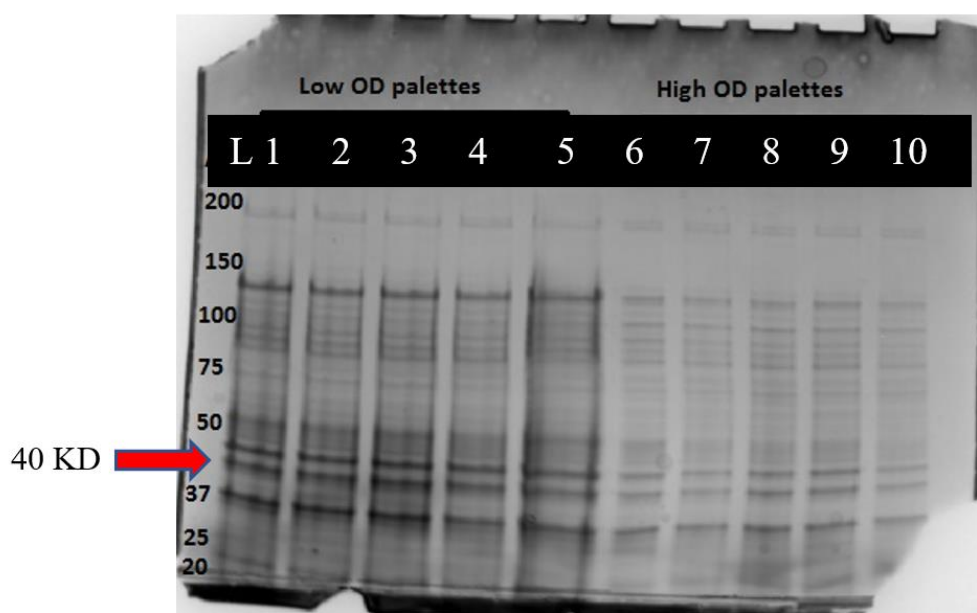


Figure 5-25: SDS-PAGE result 1. L: up to 200 KB protein ladder. Low OD pellets of Lane 1: pET; Lane 2: *O. basilicum* (EOMT); Lane 3: *O. basilicum* (CVOMT); Lane 4: *O. tenuiflorum*; Lane 5: *O. gratissimum*. High OD palate of Lane 6: pET; Lane 7: *O. basilicum* (EOMT); Lane 8: *O. basilicum* (CVOMT); Lane 9: *O. tenuiflorum*; Lane 10: *O. gratissimum*.

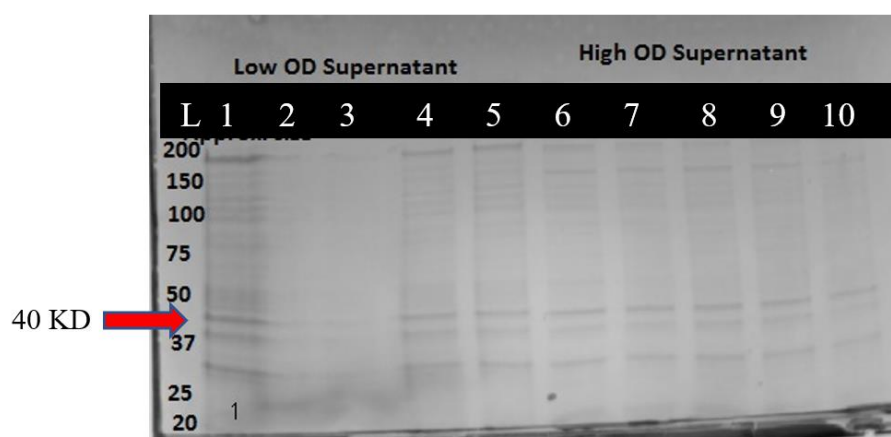


Figure 5-26: SDS-PAGE result 2. L: up to 200 KB protein ladder. Low OD supernatant of Lane 1: pET; Lane 2: *O. basilicum* (EOMT); Lane 3: *O. basilicum* (CVOMT); Lane 4: *O. tenuiflorum*; Lane 5: *O. gratissimum*. High OD supernatant of Lane 6: pET; Lane 7: *O. basilicum* (EOMT); Lane 8: *O. basilicum* (CVOMT); Lane 9: *O. tenuiflorum*; Lane 10: *O. gratissimum*

5.4 Discussion

5.4.1 RNA extraction and cDNA synthesis of *O. tenuiflorum* plant sample

Total RNA was isolated from the two *Ocimum* species, *O. tenuiflorum* and *O. basilicum*. The A260 and A280 ratios were found to be quite low, indicating low quality RNA samples. RNA samples were checked by adding loading dye at a 1:5 ratio with a 1% w/v agarose gel. These RNA extraction samples together with the c-DNA products and the RNA samples [Figure 5-9 (b)] revealed the presence of genomic DNA contamination in the RNA samples. Turbo-DNA clean-up was therefore carried out to obtain a clean RNA product.

As shown in Figure 5-10, a comparison of the genomic DNA in Lane 1 with the RNA sample in Lane 2 showed that the band was not present in the former lane, proving that genomic DNA contamination had successfully been removed from the RNA sample that was subsequently used for cDNA synthesis. The band for the c-DNA product can be seen in Lane 3. The cDNA synthesis kit provides a novel approach to synthesising cDNA using a PCR-based technique. The final product is efficient and free from genomic DNA traces.

The first strand of cDNA was synthesised from the clean RNA sample followed by an amplification process using a specific cDNA synthesis kit. As shown in Figure 5-10, distinct unsmeared bands appear for the *O. tenuiflorum* sample with an approximate length of 1,100bp on gel, the size expected. These gel bands were later excised for further procedures.

5.4.2 Obtaining of OMT sequences for *O. basilicum*, *O. tenuiflorum* and *O. gratissimum*

There were several *OMT* nucleotide sequences of *Ocimum* species present on the online databases, including *O. basilicum EOMT* and *CVOMT* sequences. These were studied in detail, and were used to design the G-blocks sequences.

Apart from *O. basilicum*, two more sequences of *O. tenuiflorum* and *O. gratissimum* were obtained from a big alignment (from Chapter 4), prepared as a part of this project. Further details of this alignment can be found in Chapter 4, “DNA and AA sequence analysis”. The two DNA sequences *O. tenuiflorum* and *O. gratissimum* were used to design another two G-blocks for *OMT* gene study.

As described in Anand *et al.* (2016), in *O. tenuiflorum* the majority of eugenol converts into methyl eugenol (ME), while in *O. gratissimum* an abundant amount of eugenol was found that does not convert into ME. As described in this study, *O. basilicum* has showed an interesting activity by producing both methyl derivatives, methyl eugenol and estragole, which are derived from eugenol and chavicol respectively. This study finds that to be the main reason for selecting different *Ocimum* sequences to design different G-blocks. By designing G-blocks for this research study and by inserting these sequences into *E. coli*, their protein properties can be studied in detail.

5.4.3 Designing a full-length cDNA sequence representing OMT gene

The four different *Ocimum* species’ DNA sequences were aligned to study the *OMT* gene and to design primer pairs for its isolation. This alignment helped to study the AA conversion in the different *Ocimum* species. The differences can be seen within the aligned *O. tenuiflorum*, *O. gratissimum*, *O. basilicum (EOMT)* and *O. basilicum (CVOMT)* AA sequences. Based on these findings, these sequences were used to design a G-block to clone the *OMT* gene in the bacteria and to study the *EOMT* and *CVOMT* proteins using SDS-PAGE.

Three pairs of primers were designed for this study. Apart from these, vector-specific primers were designed for colony screening and other purposes during the study of bacterial expression. These primers were used for the PCR test as well as obtaining sequencing results using miniprep plasmid samples.

External (*EOMT*) and internal primer pairs (as discussed in Chapter 4) covered the majority of the *OMT* region, yet they were unable to cover the beginning and end of the sequence. Interestingly, published *O. basilicum EOMT* and *O. basilicum CVOMT* sequences differ slightly by a few base pairs at the beginning, although they are similar at the end. *EOMT*-F-1 and *CVOMT*-F-1 primers were therefore designed to identify two different regions, while only one reverse *OMT*-R-1074 primer was designed. By using this specially designed primer pair, a full length *OMT* sequence has been successfully isolated with the help of *EOMT* and internal primers.

EOMT and M13 primers were used to screen *pGEM* colonies, while *pET*, *EOMT* and T7 primers were used for *pET* colony screening.

A number of primers were designed and used after studying the GOI in combination with the *pET* vector sequence. IDT PrimerQuest is a useful tool to identify relevant regions on genes so as to design primers according to individual requirements. This software recommends using common sequences for different species in the same genus. A huge alignment was thus prepared, including sequences of different species sequences, a common consensus sequence being extracted from it to design primers that were used for all sample types. All primers were labelled using their genus name and vector combination. For example, *Ocimum* primers covering RE and poly A tails are parts of Gblocks and are identified as '*pET_EOMT_F*' and '*pET_CVOMT_F*' (forward primers covering initial part of *EOMT* and *CVOMT* regions accordingly). As termination regions are similar for both *EOMT* and *CVOMT* regions, a common reverse primer (*pET_OMT_R*) was designed.

The process by which the GOI was inserted into the suitable vector was carried out by creating the same restriction enzyme sites on both gene and vector. For this project, a strategy of creating two restriction enzyme sites at each end was designed in order to insert the GOI in the right orientation. For *pGEM* insertion *NotI* and *EcoRI* restriction sites were used, while for *pET* insertion *NheI* and *EcoRI* sites were added to the 3' and 5' end respectively. A *pET*-specific primer pair was designed for this replacement procedure. As G-Blocks contain two restriction sites (highlighted below), the latter were also added using primer design along with the Poly-A tail and the short part of the sequence itself.

<i>pET_EOMT-F</i>	- AAAAAAAGCTAGCATGGCATTGCAAAAAGTA
<i>pET_CVOMT-F</i>	- AAAAAAAGCTAGCATGGCATTGCAAAATATG
<i>pET_OMT-R</i>	- TTTTTTGAATTCCTTAAGGATAAGCCTCTAT

5.4.4 Colony screening sequences

Screening colonies using PCR while checking the reliability of the results is complicated. After the transformation process, non-transformed bacteria remain on the plate along with transformed ones. Therefore, while selecting colonies for PCR, untransformed DNA are also used as a template. Preparation of the miniprep product and confirmation of the presence of the GOI by checking their sequencing results is therefore recommended.

For *pGEM* insertion, the blue-white screening method was used. When the GOI interrupted the *lacZα* sequence as the LB plates (prepared using X-Gal), white colonies appeared, while non-transformed colonies produced a blue colour. As white colonies could also have false results, further PCR-based screening was carried out. After transformation, different amounts of the ligated mixture (100µl, 150µl, 200µl and a concentrated solution) were plated on LB plates prepared with X-gal and IPTG. After overnight incubation, non-recombinant and recombinant colonies appeared as blue and white respectively. A synthesis of β-galactosidase was disrupted in recombinant colonies and therefore they were unable to produce the blue colour, while non-recombinant colonies had functional β-galactosidase and were thus able to do so. The number of colonies increased with the amount of ligation mixture. The number of white colonies appeared on LB plates, which were further plated to check for false results; eight out of 10 colonies turned blue with overnight incubation. Colonies that remained white after the second incubation period were only used for further processes.

As shown in Figure 5-12, the fourth and fourteenth colonies containing the *O. tenuiflorum* cDNA sequence were selected out of the total 14 colonies. These two colonies were later used to inoculate the LB culture overnight. The mini prep plasmid DNA samples were prepared to check the presence and orientation of the GOI. The BLAST search showed the presence of the *EOMT* gene in both the fourth and fourteenth colonies in Figure 5-13, while Figure 5-14 shows the correct orientation of the *EOMT* gene. The initial gene region containing the ATG start codon was annotated in the chromatograph of these two colonies' sequencing results.

Unfortunately, the same results could not be achieved for *O. gratissimum* and *O. basilicum* cDNA sequences. The remaining experiment was therefore designed by using G-Blocks containing four sequences with the *pET* vector.

For *pET* colony screening, after obtaining positive results with *pET* primers using the G-block as a template, either PCR-purified samples in cases of faint bands or gel-purified products while obtaining single, sharp bands on the gel were used as a template for the ligation process. Around 1,100bp gene fragment was ligated to the *pET* vector to transform the electrocompetent *E. coli* cells. Ligation products were confirmed through the PCR reaction. After transformation, various amounts of the ligated mixture were plated on LB plates prepared with kanamycin. The colony numbers had increased with the amount of the ligation mixture.

After growing colonies overnight, recombinant colonies were screened to determine whether the GOI was present. Plasmid DNAs were prepared from selected colonies. The colony PCR technique was later developed for quick identification, as plasmid DNA preparation takes one more day to grow the culture.

Figure 5-20 PCR test confirms the presence of four selected colonies transformed by the *pET* vector and relevant G-blocks. The second colony of *O. basilicum* (*EOMT*), the 23rd of *O. basilicum* (*CVOMT*), the third of *O. tenuiflorum* and the first of *O. gratissimum* were selected for mini-prep plasmid preparation and were used to check the presence of the GOI and orientation through sequencing results. The chromatogram results shown in Figure 5-21, Figure 5-22, Figure 5-23 and Figure 5-24 illustrated the annotated initial part of the gene containing ATG start codon. These results not only confirm the presence of the GOI in various colonies (the *EOMT* gene in the second colony of *O. basilicum* (*EOMT*), the third colony of *O. tenuiflorum* and the first of *O. gratissimum*, as well as the *CVOMT* gene in the 23rd colony of *O. basilicum* (*CVOMT*), but also confirmed the correct orientation of the *OMT* gene. All four colonies were later used for IPTG induction followed by SDS-PAGE analysis.

5.4.5 Summary of vector cloning results output

The work described in this chapter described the cloning of a randomly fragmented genome of *O-methyltransferase* (1,074bp) in a manner that at genome insert scale could be transferred to

another vector. An ORF selection of the *OMT* gene for eugenol and chavicol was found to be useful in selecting a linear fragment. This selected insert into the expression vector followed by a PCR-based colony screening can create libraries of a reasonable size. Two types of cDNA sequences representing *EOMT* and *CVOMT* genes were found to be almost identical. There were different *Ocimum* species, and chemotypes were used to study the activity of the targeted phenylpropanoids. *OMT* gene analysis clearly showed the highest amount of phenylpropanoid levels in young tissues (plant, inflorescence) within the *Ocimum* genus. In particular, this project has extensively studied methyl eugenol levels within different chemotypes of *O. tenuiflorum*. Further study characterised *O. basilicum* and *O. gratissimum* species varying in their phenylpropanoid contents. Moreover, a study demonstrated that *EOMT* and *CVOMT* enzymes displays their highest activity on preferred substrates, eugenol and chavicol respectively, although they managed to produce some amounts of methyl derivatives while acting on non-preferred substrates. In other words, *EOMT* can produce some estragole while using chavicol as a substrate, while *CVOMT* can produce some methyl eugenol by using eugenol as a substrate. This study compared two identical genes, *EOMT* and *CVOMT*, using cloning strategies within different *Ocimum* species.

5.4.6 Protein production study

The protein production study was begun by using IPTG induction followed by SDS-PAGE Assay, in which known sample protein presence are checked. Unknown protein samples and dilutions are treated in the same manner to measure their absorbance levels using a spectrophotometer. However, shortages of time and the unavailability of certain equipment and materials complicated an interpretation of the result. Some other factors, including repetition of replicates, correction of blank readings and dilution steps, play an important part in obtaining fruitful results. This situation is responsible for the confusing calculations required to arrive at the final determination.

5.5 Conclusion

The objective of this chapter has been to isolate and characterise a full length *OMT* gene from *O. tenuiflorum*, *O. gratissimum* and *O. basilicum*. Evidence from molecular and phylogenetic analyses shows that *EOMT* and *CVOMT* sequences have been successfully isolated from *O. basilicum* (Gang *et al.*, 2001; Gang *et al.*, 2002). Since both cloned and sequenced *EOMT* and *CVOMT* sequences represented the two different sequences, it can be

hypothesised that two genes, *EOMT* and *CVOMT*, are present in *O. basilicum*. Additionally, Gang *et al.*, 2002 managed to confirm by site-directed mutagenesis that the single AA difference (*EOMT* (S261) and *CVOMT* (F260)) appears to be responsible for the substrate preference of these enzymes (Gang *et al.*, 2002).

EOMT was shown to have an enhanced expression during pre-flowering, but decreased at flowering and further at post-flowering. These findings may also be related to the amount of eugenol and its methyl derivative, methyl eugenol, at various developmental stages within *O. tenuiflorum* chemotypes (Renu *et al.*, 2014).

Experiments in comparative studies have shown that eugenol and chavicol, like phenylpropanoids, are one of the major secondary metabolites within the genus *Ocimum*. It is also suggested that the gene from one species can play an important role in the same pathway in another, but their function may change (Anand *et al.*, 2016).

The activity of *OMT* enzymes from phenylpropanoid pathways can modulate the levels of certain secondary metabolites within different *Ocimum* species. The study of the *OMT* gene shows that they play a role in accordance with the varied accumulation of eugenol and chavicol levels in the genus *Ocimum* (Renu *et al.*, 2014; Anand *et al.* 2016). Based on a combination of these previous findings, the *EOMT* gene from *O. tenuiflorum* and *O. gratissimum*; while both *EOMT* and *CVOMT* genes from *O. basilicum* have been successfully isolated and cloned in *E. coli* using *pGEM* and *pET* expression vectors. The presence of *EOMT* and *CVOMT* enzymes in *O. basilicum* has already been confirmed by Gang (Gnag *et al.*, 2002), but the novelty of this project regards the activity of *CVOMT* and the resulting absence of chavicol and estragole in *O. tenuiflorum* and *O. gratissimum*. This can be hypothesised from the PCR results using *EOMT*-F-1 and *CVOMT*-F-1, where bands were absent while using *O. tenuiflorum* and *O. gratissimum* genomic DNA samples as a template. It can be concluded from these results either that only one enzyme is responsible for the production of both methyl derivatives (methyl eugenol and estragole) depending on substrate availability within each *Ocimum* species, or two enzymes are expressed in certain species but not in others. In other words, the *EOMT* enzyme was found to be active in both *O. tenuiflorum* and *O. gratissimum*, while the *CVOMT* enzyme did not express at all in these two species. As a result, no traces of chavicol or estragole were found in these two species.

5.5.1 Further work

This chapter results could be further explained by the successful plant transformation study by detecting the presence of two separate *OMT* genes that are expressed at the different level in different *Ocimum* species. The study approach can offer additional insights into some other factors that could modulate *EOMT* and *CVOMT* activities.

6 *O*-methyltransferase binary vector preparation for the cloning of novel plant transformation constructs

6.1 Introduction

An *Agrobacterium* contains a large Ti plasmid that can produce tumour-like structures on plants. It can transfer T-DNA into the plant's genome. This plasmid has a growth regulator and opine synthesis genes. It can act on a wide range of plants including dicotyledon and gymnosperm. This natural plant gene transfer techniques have been widely used to transform plants (Zambryski *et al.*, 1980). Several modifications require prior to the use of Ti plasmids for the transformation process with regard to the removal of tumour-inducing genes and the addition of markers to identify genes after transformation (Garfinkel *et al.*, 1981). The whole process of inserting foreign genes into Ti plasmids has been complicated until the introduction of binary vectors during the plant transformation process. The binary system contains binary and helper plasmids. It is also known as the 'Shuttle vector' as it can replicate in different organisms including *E. coli* and *Agrobacterium* (Bevan, 1984).

6.1.1 Different gene transfer methods in plants

Gene transfer methods in plants fall into two categories: vector-based or indirect gene transfer and vectorless or direct gene transfer. Among the former, the *Agrobacterium tumefaciens* containing the Ti plasmid has been widely used. The direct gene transfer methods are microinjection, chemical mediation, electroporation, particle bombardment, transformation using bacteria, lipofection and conjunction.

6.1.2 Plant transformation

A process of inserting newly designed genes into plant chromosomes is called 'plant transformation'. This process uses a wide range of tissue culture methods. Plant regeneration is the key step during transformation. Several options are available to insert GOIs into plant chromosomes, *Agrobacterium*-mediated transformation being a hugely popular method. This process can dramatically change the plant's characteristics to improve the quality of human

life (Khan *et al.*, 2016). By using a variety of tools such as selection markers, the transformed and non-transformed plants can be easily identified, which is also helpful when comparing control and experiment tissues.

Arabidopsis has been used as a model plant organism during the plant transformation process. *C. elegans* (The *C. elegans* Sequencing Consortium, 1998) and *Drosophila* (Adams, 2000) genome structures had been studied, but they provided a very limited amount of information regarding physiological and evolutionary processes. The short life cycles, small genomes and multitude of flowerings are key aspects of the *Arabidopsis thaliana* plant that make it ideal for genome study. Due to its genomic variation, it is widely popular in molecular biology, genetics biochemistry and many other fields as a model plant (Meyerowitz, 1987). A large number of protein kinase receptors have been identified from this plant, among which more than 25 proteins are studied in detail to discover their biological functions (Abrc, 2012), which include CLV1-differentiated meristem cells between floral and shoot formation (Clark *et al.*, 1997), VH1-effected differentiation in leaf patterns (Clay & Nelson, 2002), ER-influenced inflorescence structure in flowers (Torii *et al.*, 1996) and the involvement of ERECTA against pathogen protection along with plant growth (Godiard *et al.*, 2003). A huge DNA library is available for mutant *Arabidopsis* genes that can be inserted into target plants to study insertion mutation (Prescott *et al.*, 2010).

There are several ways of obtaining or designing the foreign genes that must be inserted into the targeted plant tissues. The common method of designing GOIs is by using various organisms such as a variety of bacteria (*E. coli*, *Agrobacterium*), yeast cells and plant species. Stable insertion and successful regeneration of second or third plant generation are two key aspects during the plant transformation process. While using *Agrobacterium* for this process, the ‘flower dip’ procedure has been applied where the plants (especially *Arabidopsis*) containing flowers are inoculated in the mixture of bacteria. After treating the targeted plants with bacteria, they are kept separated from the others in order to collect the seeds and study them for the success of the next generation of plant transformation. This method’s ease means that it can also be applied to other plant species, though the success rate for *Arabidopsis thaliana* is the greatest.

6.1.3 Plants transformed with homologues of the OMT gene in various plants

Apart from eugenol, chavicol and their methyl derivatives, other phenylpropenes including iso eugenol, isochavicol, trans-Anethole, methylisoeugenol and isoestragole (Obeng-Ofori and Reichmuth, 1997) are produced by plants of basil, cloves, star anise and fennel (Figure 6-1) (Orav *et al.*, 2008).

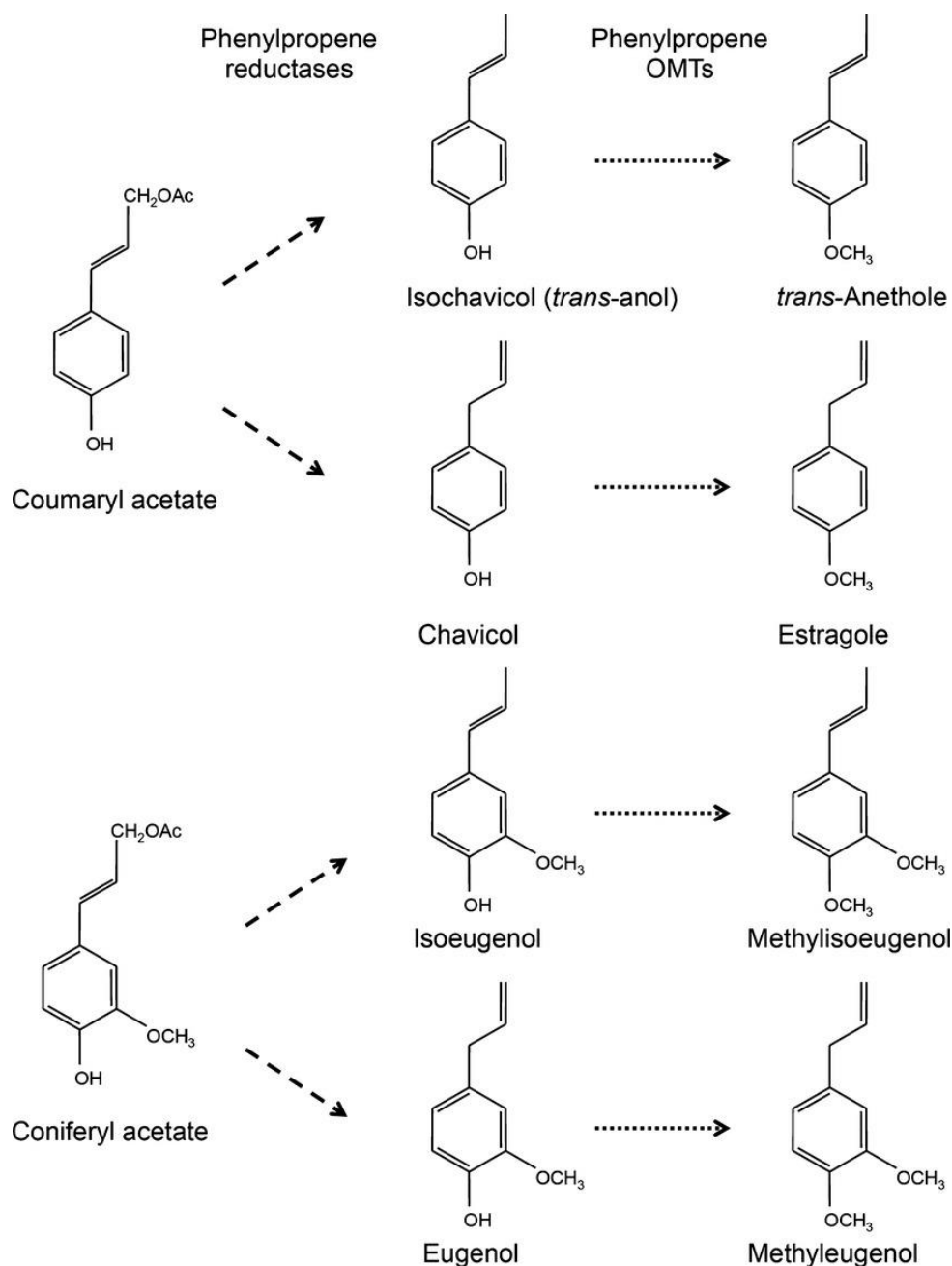


Figure 6-1: Different phenylpropene production in plants. The reaction catalysed by *O*-methyl transferases (dotted lines) and NADPH-dependent phenylpropene reductases (dashed lines) (Yauk *et al.*, 2015)

According to Yauk *et al.* (2015), an estragole might contribute the unique flavours in some varieties of apple. A bioinformatics approach was used to identify the gene responsible for

estragole production; its kinetic study confirmed that the specific *MdoOMT1* allele is responsible for estragole production. The same research group later managed to produce the transgenic plants which confirmed the gene's activity directed at estragole production and its role in the apple plant (Yauk *et al.*, 2015). Some other *OMT* genes including the isoflavone 3'-*O*-methyltransferase from *Pueraria lobata* roots (Li *et al.*, 2014) and caffeic acid 3-*O*-methyltransferase from wheat (Wang *et al.*, 2018) have also been isolated.

6.1.4 Successful transformation of *Ocimum* plant

The first transgenic *Ocimum* plant was produced using *Agrobacterium* strains GV3101 and EHA105 where *O. basilicum* and *O. citriodorum* species were successfully regenerated with the β -Glucuronidase (GUS) gene (Deschamps and Simon, 2002). Since then several research projects transformed several genes into the various parts of the *Ocimum* plant. Leaves as well as hypocotyls and excised shoots were used for invitro germination, and the efficiency of transformation was found to be highest in the leaves of *O. tenuiflorum* (Vyas and Mukhopadhyay, 2014). The concentrations of MS medium along with various growth hormones were used to focus on the development of an efficient transformation protocol in *O. gratissimum* (Khan *et al.*, 2015).

6.2 Material and method

As described in Chapter 5 (Section 5.3.3), *OMT* sequences (both *EOMT* and *CVOMT*) were used to design four G-blocks. These G-blocks were used as a template in PCR to add *Att* sites at the end of each G-block fragment for Gateway cloning.

6.2.1 Gateway pDONR™221 vector (to create entry clone)

Gateway cloning vectors are specially designed with '*Att*' sites which can easily be flanked along with GOI/G-blocks. This technique is more convenient than the use of restriction enzymes. A successful entry clone can easily be shuttled into the majority of expression vectors (Figure 6-2). With the help of the manufacturer's instructions and Gateway® LR Clonase® II (Invitrogen, Cat. No.11791-043 USA), the product was then transferred to the binary vector.

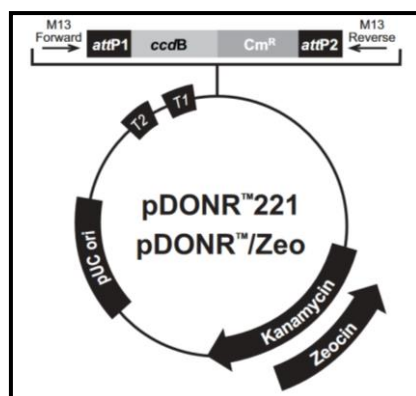


Figure 6-2: *pDONR™221* vector map

6.2.1.1 *Att* primers

After isolating the plasmid DNA and while attaching it to the Gateway cloning vectors, *Att* sites (highlighted in red) along with the ‘Shine-Dalgarno (SD) sequence’ (highlighted in yellow) were added to design these specific primers:

1. *Att*_EOMT-F 5'-**GGGGACAAGTTTGTACAAAAAAGCAGGCT****AAAAA**ATGGCATTGCAAAAAGTAGATATTT-3'
2. *Att*_CVOMT-F 5'-**GGGGACAAGTTTGTACAAAAAAGCAGGCT****AAAAA**ATGGCATTGCAAAAATATGGATATTT-3'
3. *Att*_OMT-R 5'-**GGGGACCACTTTGTACAAGAAAGCTGGGT**TTAAGGATAAGCCTCTATGAGAGAC-3'

All PCR reactions were carried out in the G-Storm GS1 Thermal Cycler and MJ Research PTC-200 Thermal Cycler machines, while all gel results were analysed using the Gel Doc EZ system.

6.2.1.2 Preparing G-blocks

The readymade G-blocks were treated according to the manufacturer’s guidelines (gBlocks® Gene Fragments, IDT) before preparing a series of dilutions for PCR as a template. 1/100, 1/500, 1/1000 and 1/5000 dilutions were tested while they were used with the *Att* primers. The complete PCR program for *Att* primers was as follows (Table 6-1):

Table 6-1: The thermocycler program for *Att* primers

<i>Att</i> primers PCR program	Thermocycler conditions
Thermocycler programme	An initial denaturation of 2 minutes at 94°C, 35 cycles of denaturation (94°C for 30 seconds), annealing (58°C for 30 seconds) and extension (72°C for 45 seconds); final extension 72°C for 2 minutes

6.2.2 Sterilisation process

Leaf segments and seeds were sterilised before establishing them on the MS medium. The material was thoroughly washed under running tap water for ten to fifteen minutes. 20% Savlon and a few drops of tween mixture were used to remove fungal infection and other debris. The material was soaked in this mixture for five to seven minutes. Autoclaved (The Prestige Medical autoclave, The Classic 2100 range) double-distilled water was used eight to ten times to remove the above mentioned mixture traces from the plant material, which was then thoroughly dried with tissue or filter paper.

Different types of *Arabidopsis thaliana* seeds were ordered from The Nottingham Arabidopsis Stock Centre (NASC) (<http://arabidopsis.info/>).

1. N1092 – Columbia wild type (Col-O)
2. N502373 – Columbia wild type with T-DNA insertion (OMT)
3. N646526 – OMT mutant

Wild-type Arabidopsis seeds were collected from the Plant Biotechnology Department of Warwick University.

Ocimum seeds were ordered from 'Jungle seeds' (www.jungleseeds.co.uk) and 'Chilton seeds' (<https://www.chilternseeds.co.uk>). Some seeds were also collected from the Hindu community in Leicester and India.

6.2.3 Medium recipe

Murashige and Skoog's (MS) medium (Cat. No. - MD154, Phygenera) was prepared by adding various organic and inorganic components. 4% sugar (Cat. No. - S0389 Sigma Sucrose) was added while preparing the MS medium. The pH was adjusted (around 5.7) using HCl or NaOH. The solid MS medium was prepared by adding 8% agar in either magenta jars or petri dishes. Various growth hormones including 0.5mg/L kinetin (Cat. No. - K0753 Sigma Kinetin plant cell culture tested, crystalline), 3mg/L picloram (Cat. No. - P5575 Sigma Picloram plant cell culture tested, BioReagent), 2mg/L BAP (Cat. No. - B3408 Sigma 6-Benzylaminopurine plant cell culture tested), 1mg/L NAA (Cat. No. - N0640 Sigma 1-Naphthaleneacetic acid plant cell culture tested, BioReagent, $\geq 95\%$, crystalline) were added to try various compositions of MS medium.

6.3 Result

While preparing the ligation for the *pDONR* and binary vectors, a unique technique called ‘Gateway Cloning’ was used. In this technique, rather than using restriction enzyme sites, specially designed ‘*Att*’ sites are added.

6.3.1 PCR purification

A Qiagen™ purification kit was used to purify the PCR product containing the G-block attached with *Att* sites. Figure 6-3 shows that it was possible to obtain the full length of the *OMT* gene from *O. basilicum* (*EOMT*) (Lane 1), *O. basilicum* (*CVOMT*) (Lane 2), *O. tenuiflorum* (Lane 3) and *O. gratissimum* (Lane 4) using *Att* primers. No bands were visible in the negative control (Lane 5). The expected product sizes were present in the gel picture.

All these PCR purified products were further used as a template during the Gateway cloning process.

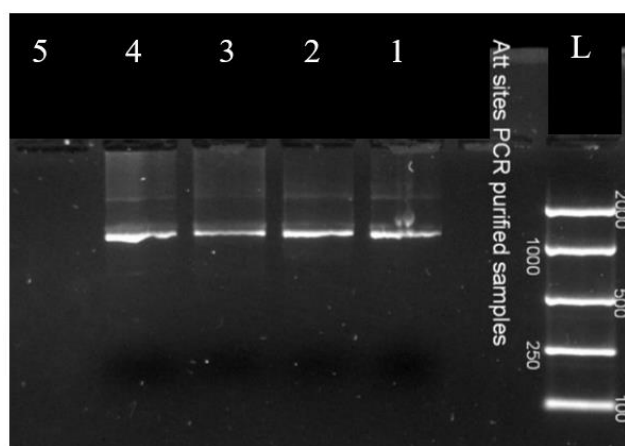


Figure 6-3: PCR-purified product containing *Att* sites. L: 2 Kb DNA ladder (EasyLadder 1, Cat. No. BIO-33045; Bioline); Lane 1: *O. basilicum* (*EOMT*); Lane 2: *O. basilicum* (*CVOMT*); Lane 3: *O. tenuiflorum*; Lane 4: *O. gratissimum*; Lane 5: negative control

6.3.2 Gateway cloning results

6.3.2.1 BP reaction

An entry clone was prepared during the BP reaction. As shown in Figure 6-4, positive colony screening has been achieved for all four G-blocks using *EOMT* primers. The selected colonies were further used to prepare mini prep plasmids for the second round of gateway cloning, which was the LR reaction.

Figure 6-4 shows that it was possible to obtain the full length of the *OMT* gene during colony screening of the BP reaction. In figure 6-4 (a), Lanes 1-8 show the band representing *O. basilicum* (*EOMT*) colonies 1 to 8 and Lanes 9-16 show the bands representing *O. basilicum* (*CVOMT*) colonies 1 to 8. In figure 6-4 (b), Lanes 1-8 show the bands representing 1 to 8 *O. tenuiflorum* colonies and lane 9-16 showed the bands representing *O. gratissimum* colonies 1 to 8 using *Att* primers. No bands were visible in the negative control. The expected product sizes were present in the gel picture. All colonies apart from the fifth, *O. basilicum* (*EOMT*), showed positive bands, from which selected colonies were sent for sequencing.

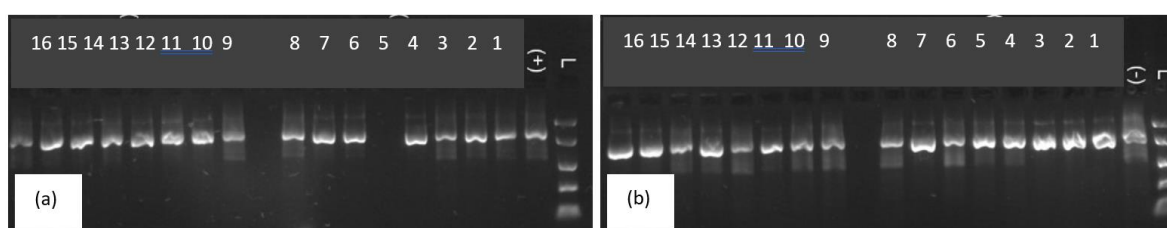


Figure 6-4: BP reaction colony screening. L: 2 Kb DNA ladder; (+): positive control; (-): negative control. (a) Lanes 1-8: *O. basilicum* (*EOMT*) colonies 1 to 8; Lanes 9-16: *O. basilicum* (*CVOMT*) colonies 1 to 8. (b) Lanes 1-8: *O. tenuiflorum* colonies 1 to 8; Lanes 9-16: *O. gratissimum* colonies 1-8

6.3.2.2 LR reaction

The GOI has been transferred from the entry clone to the destination vector during the LR reaction. The binary vector is used here as a destination vector from which the GOI can be transferred directly into the targeted plant species. As shown in Figure 6-5, positive colony screening results were achieved using *EOMT* primers. Selected colonies were further grown overnight in liquid LB to obtain mini prep plasmid DNA samples.

Figure 6-5 shows that it was possible to obtain the full length of the *OMT* gene during colony screening of the LR reaction. In figure 6-5(a), Lanes 1-8 show the bands representing *O. basilicum* (*EOMT*) colonies 1 to 8 (the 2nd, 5th and 7th colonies exhibit bands); Lanes 9-16 show the bands representing *O. basilicum* (*CVOMT*) colonies 1 to 8 (all eight colonies evince bands). In figure 6-5(b), Lanes 1-8 show the bands representing *O. tenuiflorum* colonies 1 to 8 (all seven colonies except the fourth showed the bands), and Lanes 9-16 show the bands representing *O. gratissimum* colonies 1 to 8 (only the second and third colonies show bands) using *Att* primers. No bands are visible in either the negative or the positive controls. As the

pDONR vector was used as a template where the *Att* sites were missing, the band was not visible in the positive control. The expected product sizes were present in the gel picture.

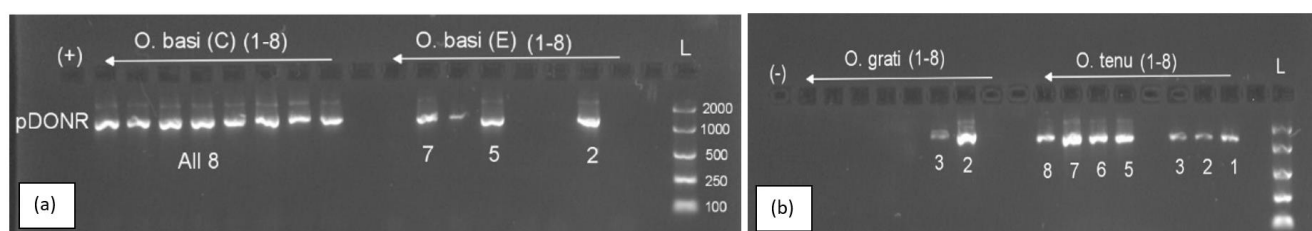


Figure 6-5: LR reaction colony screening. L: DNA ladder; (+): positive control; (-): negative control. (a) *O. basilicum* (EOMT) colonies 1-8; *O. basilicum* (CVOMT) colonies 1-8. (b) *O. tenuiflorum* colonies 1-8 and *O. gratissimum* colonies 1-8. All colonies showed positive bands apart from 1st, 3rd, 4th and 8th *O. basilicum* (EOMT), 4th *O. tenuiflorum* and 1st, 4th, 5th, 6th, 7th and 8th *O. gratissimum* colonies

6.3.2.2.1 LR colony screening

Att-F and *Att*-R primers were used to check full length sequences. The PCR was designed by using *Att*-F and Internal-R primers. The PCR was performed as described in the ‘General material and method’ chapter.

Figure 6-6 shows the colony screening conformation of the LR reaction using *Att* primers. The *Att_EOMT* PCR products were included in Lanes 3-8, where Lane 3 represents the fifth *O. basilicum* (EOMT) colony, Lane 4 the seventh *O. basilicum* (CVOMT) colony, Lanes 5 and 6 the fifth and seventh *O. tenuiflorum* colonies and Lane 7 the third *O. gratissimum* colony. No band was visible for the eighth *O. gratissimum* colony. The *Att_CVOMT* PCR products are included in Lanes 10-14, where Lane 12 and 13 represent the second and eighth *O. basilicum* (CVOMT) colonies. No bands were visible in the negative control. The expected product sizes were present in the gel picture, including positive controls in Lanes 1, 2, 10 and 11.

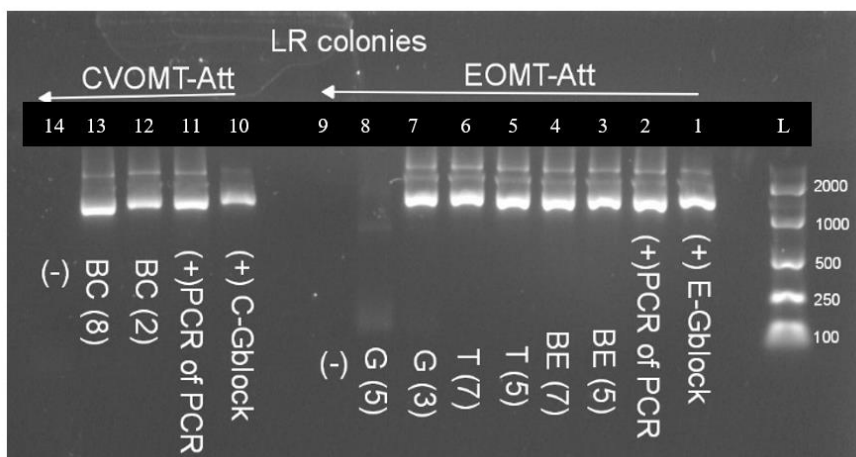


Figure 6-6: LR colony screening using EOMT *Att* primers. L: DNA ladder; Lane 1: EOMT G-block as a positive control; Lane 2: PCR of PCR as positive control; Lane 3: 5th *O. basilicum* (EOMT) colony; Lane 4: 7th *O. basilicum* (EOMT) colony; Lane 5: 5th *O. tenuiflorum* colony; Lane 6: 7th *O. tenuiflorum* colony; Lane 7, 3rd *O. gratissimum* colony; Lane 8, 5th *O. gratissimum* colony; Lane 9: negative control. LR colony screening using CVOMT *Att* primers. Lane 10: CVOMT G-block as a positive control; Lane 11: PCR of PCR as positive control; Lane 12: 2nd *O. basilicum* (CVOMT) colony; Lane 13: 8th *O. basilicum* (CVOMT) colony; Lane 14: negative control.

Figure 6-7 shows the selected colony PCR conformation from each G-block using a combination of *Att* and internal primers. The *Att_EOMT* and Int-R primer PCR products were included in figure 6-7(a), where Lane 1 represents the positive control, Lane 2 the fifth *O. basilicum* (EOMT) colony, Lane 3 the fifth *O. tenuiflorum* colony and Lane 4 the third *O. gratissimum* colony. No bands were visible in the negative control (Lane 5).

The *Att_CVOMT* and Int-R primer PCR products were included in figure 6-7(b), where Lane 1 represents the positive control and Lane 2 the second *O. basilicum* (CVOMT) colony. No bands were visible in the negative control (Lane 3). The expected product sizes were present in the gel picture, including positive controls in Lanes 1, 2, 10 and 11. As multiple bands present in the selected colonies, the correct size bands were excised, purified and sent for sequencing.

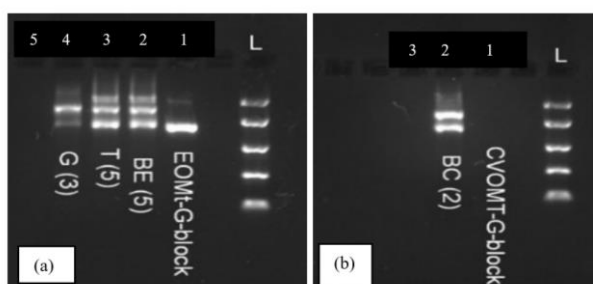


Figure 6-7: *ATT-F* primers in the combination with internal reverse (Int-R) primer. PCR with EOMT *Att-F* and Int-R primers. L: DNA ladder; Lane 1: EOMT G-block as a positive control; Lane 2: 5th *O. basilicum* (EOMT) colony; Lane 3: 5th *O. tenuiflorum* and Lane 4: 3rd *O. gratissimum* colonies; Lane 5: negative control. PCR with CVOMT *Att-F* and Int-R primers. L: DNA ladder; Lane 1: CVOMT G-block as a positive control; Lane 2, 2nd *O. basilicum* (CVOMT) colony; Lane 3: negative control.

Bioinformatics studies were also carried out for all four colonies: the fifth *O. basilicum* (EOMT), fifth of *O. tenuiflorum*, third of *O. gratissimum* and second of *O. basilicum* (CVOMT) colonies. All the BLAST search results show the GOI's presence. Later contigs were prepared by using CLC Bio Main Workbench to check the orientation of selected colonies. The *Att*-F primer was used to check the vector's presence in combination with the Internal-R primer to check for the GOI's presence and correct direction. All sequences were analysed the same way to check the orientation of the GOI (*O. basilicum* (EOMT), *O. basilicum* (CVOMT), *O. tenuiflorum* and *O. gratissimum* respectively). As shown in figure 6-8, figure 6-9, figure 6-10 and figure 6-11, the GOI's orientations were checked in the selected LR colonies.

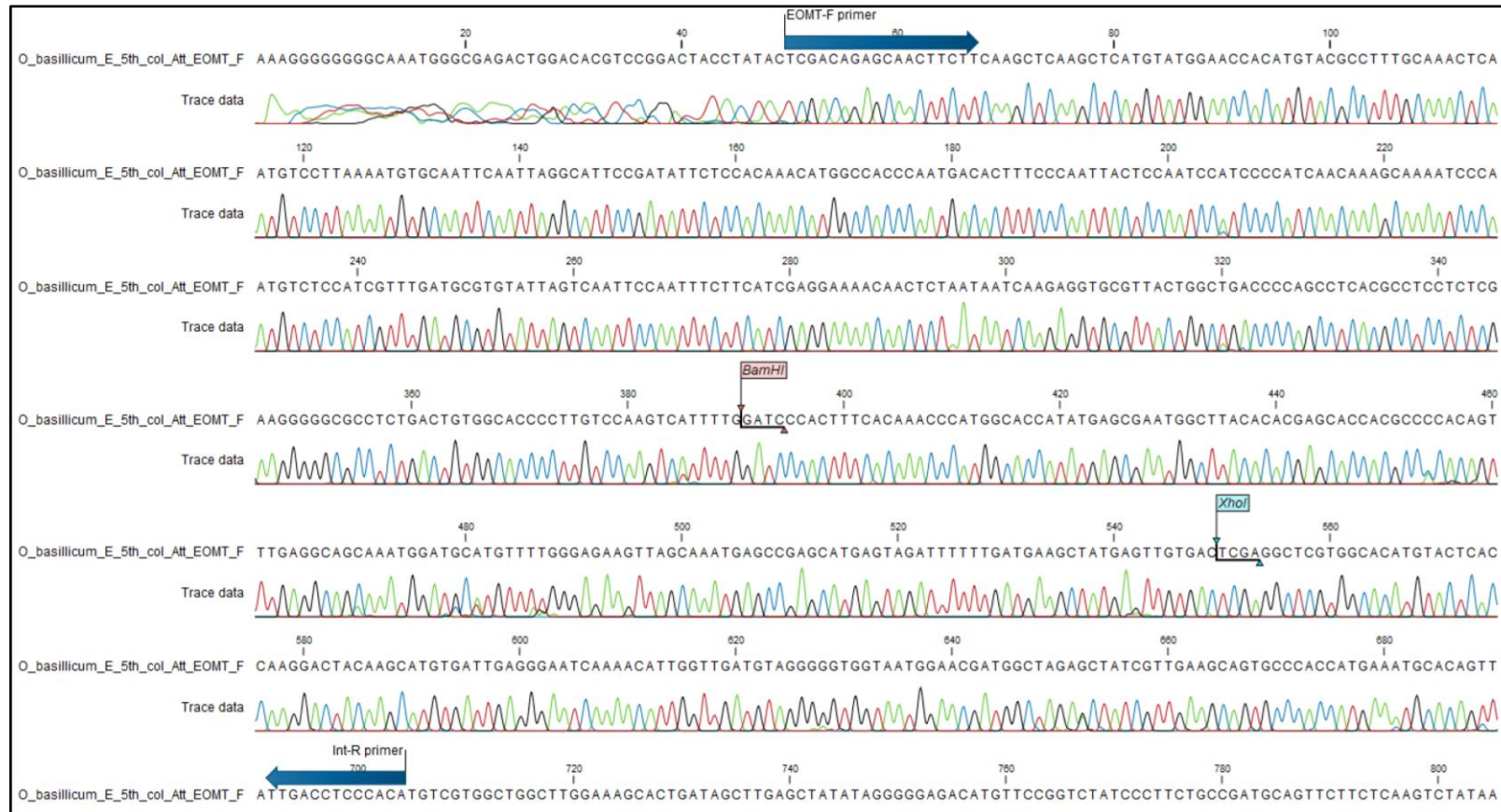


Figure 6-8: Nucleotide sequence of the PCR-amplified OMT region from the transformed 5th LR *O. basilicum* EOMT colony. Arrows show the location and orientation of the *EOMT-F* and *Int-R* primer regions. Sequencing results show external and internal regions are annotated in correct orientation.

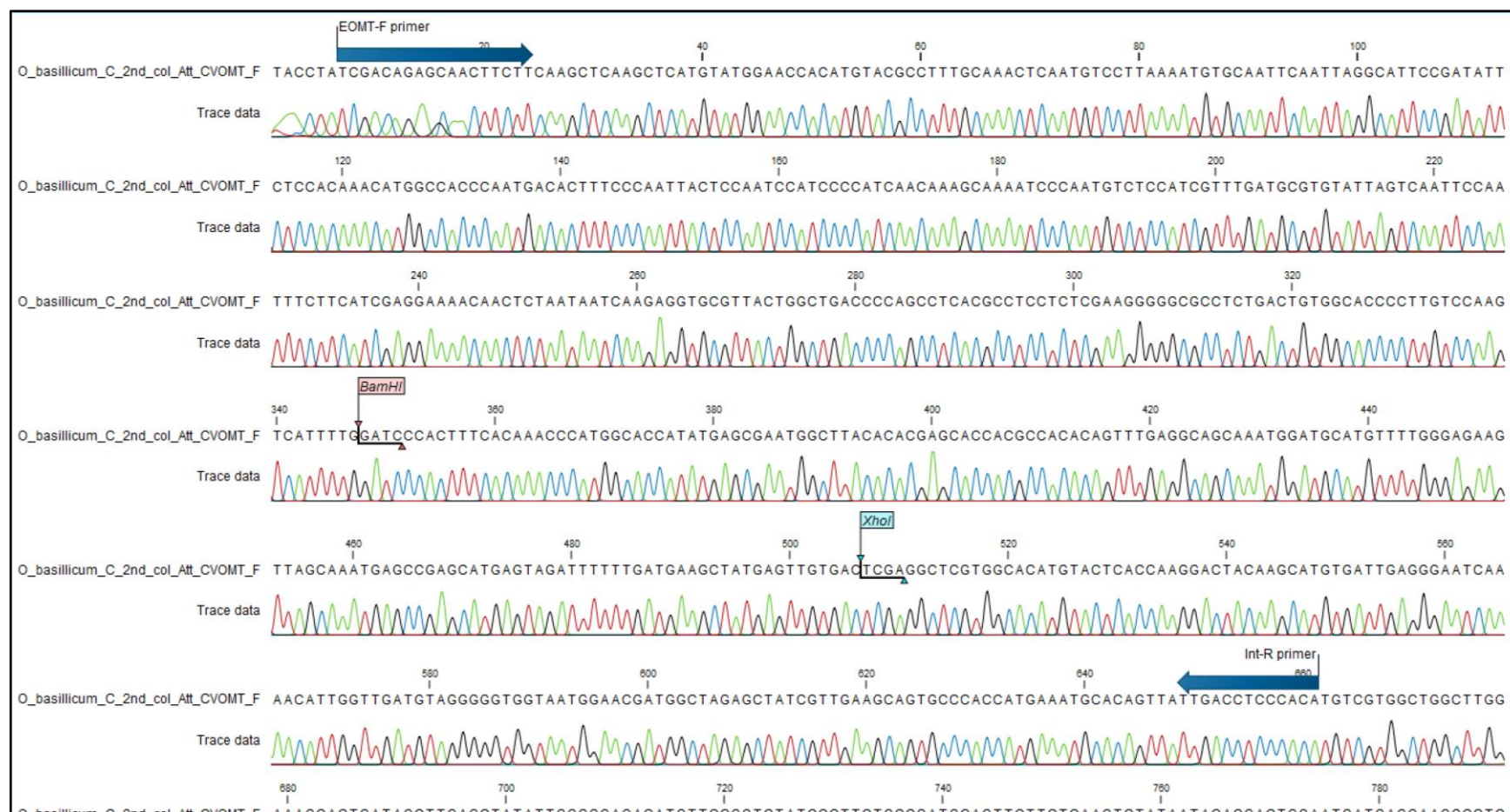


Figure 6-9: Nucleotide sequence of the PCR-amplified OMT region from the transformed 2nd LR *O. basilicum* CVOMT colony. Arrows show the location and orientation of the *EOMT-F* and *Int-R* primer regions. Sequencing results showing external and internal regions are annotated in correct orientation.

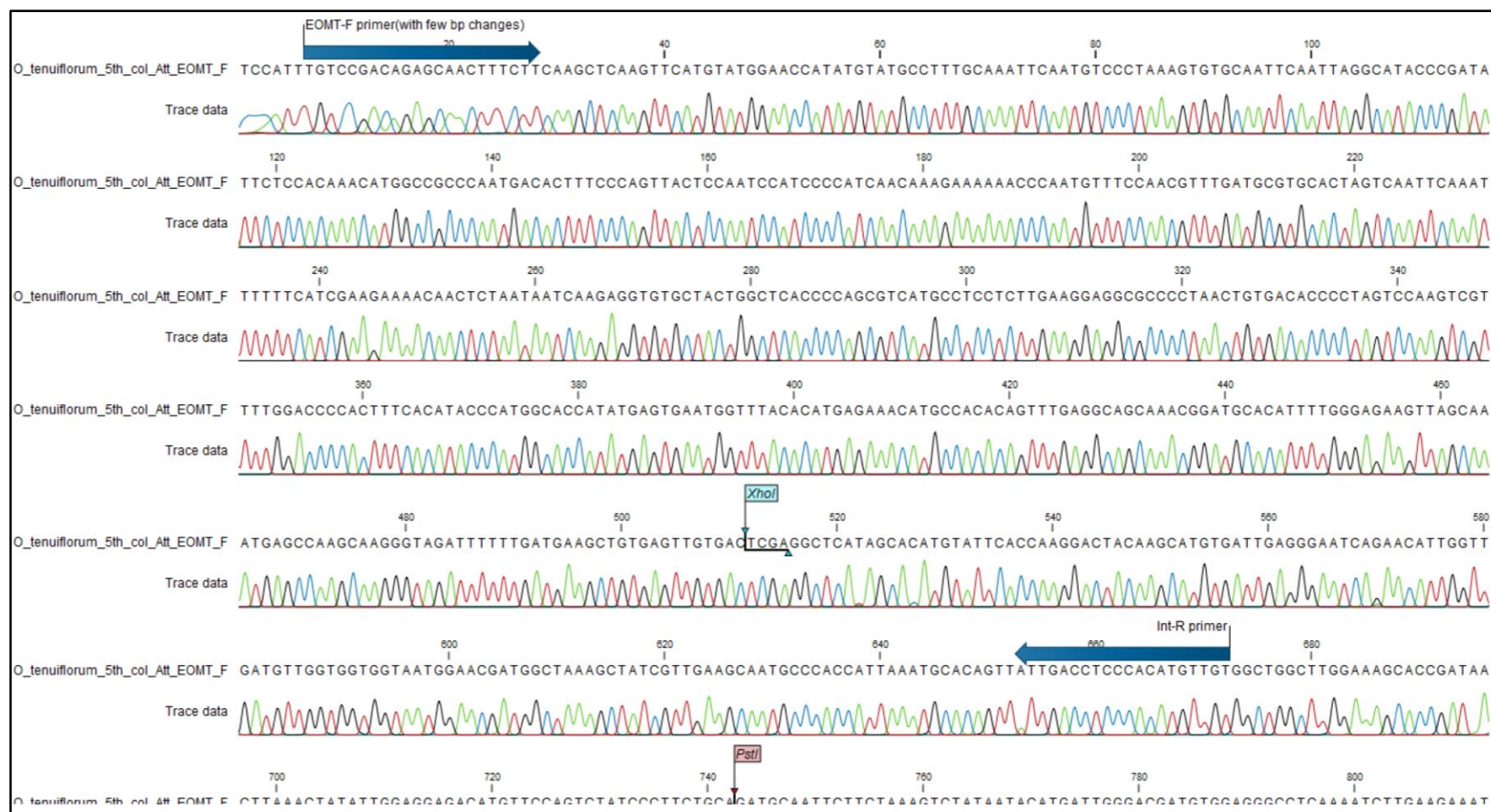


Figure 6-10: Nucleotide sequence of the PCR-amplified OMT region from the transformed 5th LR *O. tenuiflorum* colony. Arrows show the location and orientation of the *EOMT*-F (with few bp changes due to poor sequence quality in the initial region) and Int-R primer regions. Sequencing results showing external and internal regions are annotated in correct orientation.

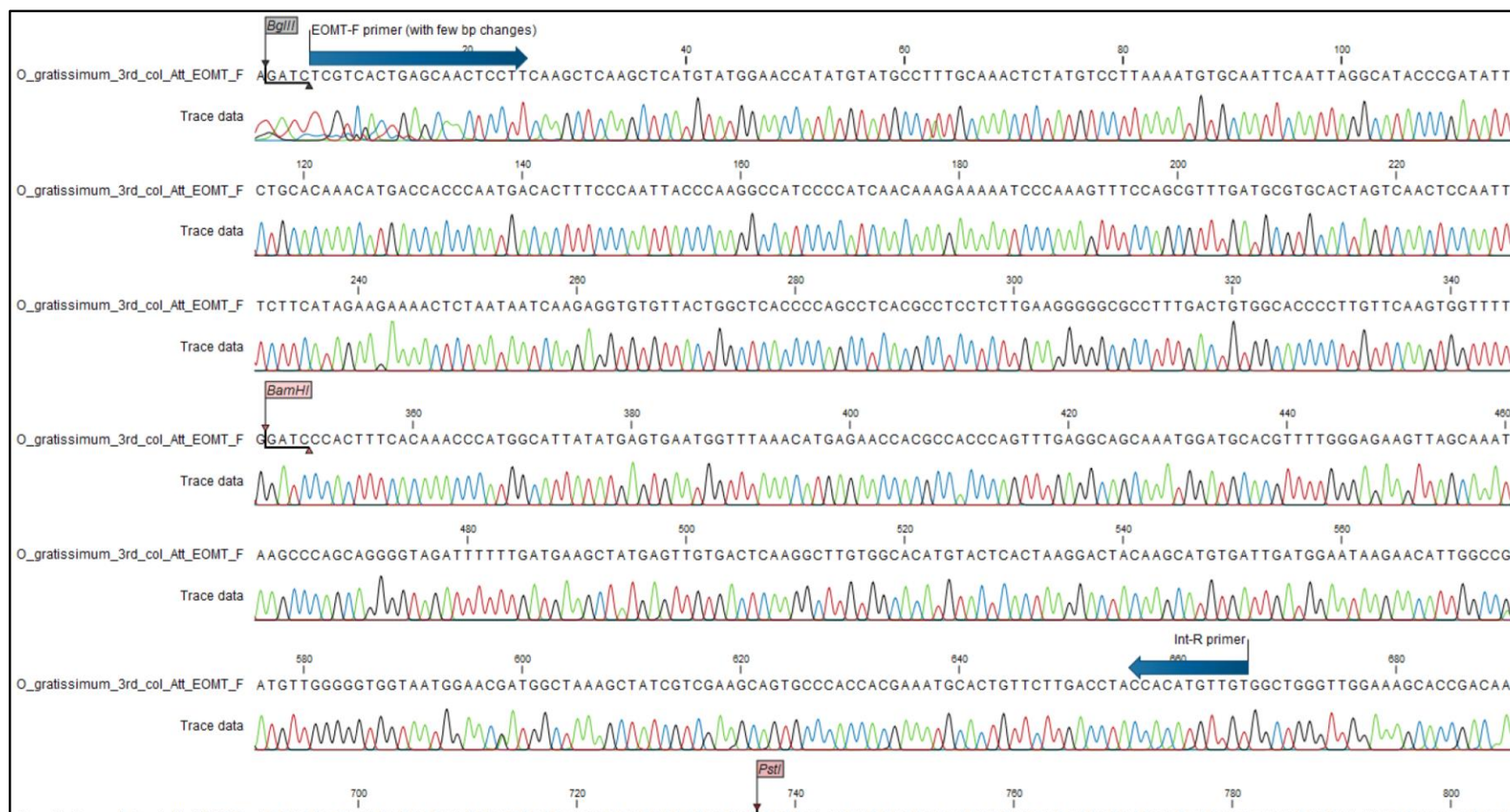


Figure 6-11: Nucleotide sequence of the PCR-amplified OMT region from the transformed 3rd LR *O. gratissimum* colony. Arrows show the location and orientation of the *EOMT*-F (with few bp changes due to poor sequence quality in initial region) and *Int-R* primer regions. Sequencing results showing external and internal regions are annotated in correct orientation.

6.3.3 Seed collection

It was necessary to check the seed dormancy at the first stage. A simple test to grow seeds on filter paper in petri dishes can help identify the seeds' activity (figure 6-12). Dormant seeds get fungal infections, while non-dormant ones can exhibit signs of fertilisation within a week.

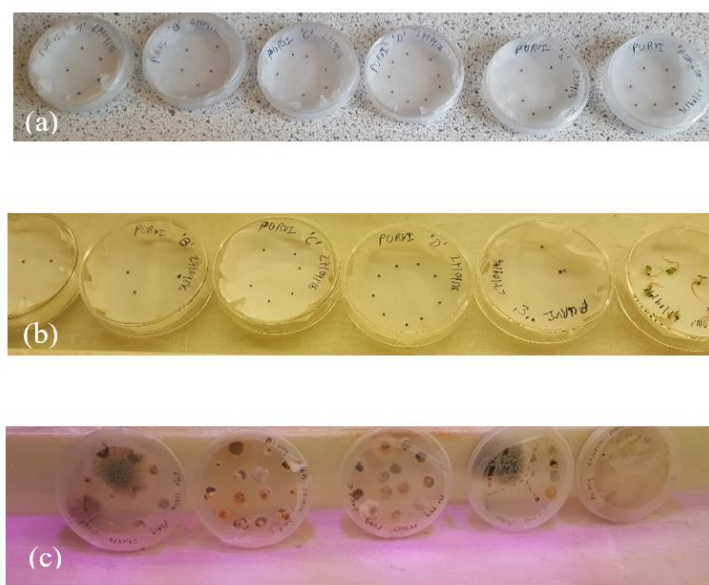


Figure 6-12: Seed germination. (a) Initial period of fertilisation (b) Germination can be seen in non-dormant seeds (c) Fungal infections on dormant seeds

6.4 Discussion

The main aim of this chapter was to introduce *EOMT* and *CVOMT* genes from different *Ocimum* species into binary vector before taking further plant transformation steps. PCR and bioinformatics results confirmed that both genes were successfully isolated from the target *Ocimum* varieties. As two different yet partly similar sequences were always derived from all result outputs, it can be suggested that two genes, *EOMT* and *CVOMT*, are present in some *Ocimum* species. Both *EOMT* and *CVOMT* genes, members of the *O-methyltransferase* gene family, act differently on the available substrates (Gang *et al.*, 2002).

Both *EOMT* and *CVOMT* gene expression levels are higher in young *Ocimum* plant parts including young leaves and inflorescence. Study reveal that these expression levels decrease as the plant reaches an advanced age. Experiments proved that same gene acted differently during the secondary metabolic pathway in different *Ocimum* species. Even though they acted on the same substrate at the same time, the quantity of the final product was different in

Ocimum varieties (Anand *et al.*, 2016). Overexpression of both genes may lead to the production of more methyl derivatives of ME and Estragole, which could be further validated by designing expression or chemical analysis of these secondary metabolites.

6.4.1 Initial stage of plant transformation

The necessity of establishing seed sterilising and germination protocols is a key finding of this chapter. Both *Arabidopsis* and *Ocimum* seeds of different species and from different suppliers or locations were obtained to initiate this experiment. As seeds were also collected locally from communities, some of them were in a dormant condition showing signs of fungal infection during their developmental stages. Henceforth, filter paper test was introduced to check for seeds' dormancy. Later, several different seed sterilisation protocols were also designed for in vitro culture. As per Dode *et al.*, 2003, 70% alcohol and sodium hypochlorite solutions were added during sterilisation process, while according to Kibler, 2014, 15% bleach solution was also tried to surface sterilise *Ocimum* seeds. Eventually, a modified version of Banu and Bari's (2007) protocol was established in which seeds were treated with 20% Savlon and a few drops of Tween-80 solution followed by three to four washes of doubly distilled water. These treated seeds were grown on an MS medium to obtain sterile in vitro shoots and root formations.

6.4.2 Gateway cloning – different ways of gene assembly

Gateway cloning technique has the primary advantage that one single recombination reaction moves a piece of DNA from one plasmid into another. This simplifies the process and reduces the time compared to restriction ligation cloning. As the binary vector used as a destination vector during this cloning, the GOI was introduced into the binary vector, which is all prepared for the next stage of the plant transformation experiment. DNA templates were designed in the form of G-blocks, which were used for PCR with *Att* primers to attach *Att* sites to the GOI, followed by ligation with the entry clone. This clone was later transferred into the binary vector. Good quality results were confirmed by sequence analysis at each step to make the whole experiment reliable.

The final product produced during the Gateway cloning process is the binary vectors containing the four GOIs. *O. basilicum* (EOMT), *O. basilicum* (CVOMT), *O. tenuiflorum* and *O. gratissimum* were prepared, which could provide a firm base for future stages of the plant

transformation process. An efficient set of these results can be used as a building block for several new plant research projects.

6.5 Conclusion

The current study has shown that the Gateway cloning system can be successfully used to prepare binary vectors for future plant transformation experiments. This represents an advantage over the various cloning techniques that do not use the RE strategy. With the use of suitable Gateway enzymes, *Att* sites were successfully flanked to prepare binary vectors containing the four G-blocks *O. basilicum* (EOMT), *O. basilicum* (CVOMT), *O. tenuiflorum* and *O. gratissimum*. The *Att* primers were designed according to Gateway cloning sites where some extra base pair sequences called ‘Shine-Dalgarno (SD) Sequences’ were added. As described by Kim *et al.* (2014), the addition of these few base pair sequences with *Att* primer sites can increase the success rates of the gene translation process.

The binary vector has become the popular choice for *Agrobacterium*-mediated genetic transformation due to its stability, ease of handling and the existence of a range of reporter genes (Ogawa *et al.*, 2008). All four G-blocks had been successfully designed and cloned into the binary vector, with impressive colony screening results.

Various seed sterilisation protocols including Kibler (2014) and Dode *et al.* (2003) were tried in order to surface sterilise *Ocimum* seeds. Eventually, a modified version of Banu and Bari’s (2007) protocol including 20% Savlon and a few drops of Tween-80 solution was implemented for *Ocimum* as well as *Arabidopsis* seeds. Together with the combination of binary vector preparation results, these protocol steps could help produce the next round of plant transformation experiment results.

Previous research findings have reported the development of an *Agrobacterium tumefaciens*-mediated genetic transformation in *O. basilicum*, *O. citriodorum* (Deschamps and Simon, 2002), *O. tenuiflorum* (Vyas and Mukhopadhyay, 2014) and *O. gratissimum* (Khan *et al.*, 2015). These publications strongly emphasise the well-developed transformation protocol for this important medicinal plant. The MS medium supplemented by various growth hormones produced the best response, with the *A. tumefaciens* strain containing the binary vector, harbouring the gus-A reporter gene, the Cauliflower Mosaic Virus (CaMV) 35S promoter and

neomycin phosphotransferase as selection marker genes to achieve the maximum transformation frequency (Deschamps and Simon, 2002; Khan *et al.*, 2015). The well-established transformation protocol could be used to investigate the novel therapeutic compound by studying its functional genomics characteristics.

6.5.1 Future work

Results achieved during this experiment, including binary vector preparation, the seed surface sterilisation technique and the MS medium protocol constitute a highly efficient set of data. These details can be further used to initiate *Ocimum* plant transformation research projects.

Two operationally different OMT enzymes, *EOMT* and *CVOMT*, were identified in a few *Ocimum* species, though separation of the relevant proteins was made difficult by the identical molecular mass (Wang and Pichersky, 1999). Two cDNA sequences for *EOMT* and *CVOMT* enzymes were later identified and used for *E. coli* expression (Gang *et al.*, 2002). An intensive study of both OMT gene expressions, their gene and encoded protein structures along with a bioinformatics-based examination would be useful to establish *Ocimum* plants as a natural source of pharmaceuticals.

The stable insertion of the transgene and the regeneration of the particular cell are both necessary for each transformation process. Selection markers and reporter genes are important tools to detect transformation sites. The plant inoculation method, in which young flowers are inoculated in bacterial culture, is famous for *Agrobacterium*-mediated plant transformation (Chang *et al.*, 1994). In order to increase transformation efficiency, some supporting techniques including particle bombardment to produce wounds at the inoculation site (Hansen & Chilton, 1996) and microinjection to insert genes into the intracellular region of bacteria (Escudero *et al.*, 1996) can also be used.

7 Chemical analysis

7.1 Introduction

In plants, a large number of secondary metabolites are biosynthesised through metabolic pathways. The majority of them are of high interest to the chemical and pharmaceutical industry. The aim of this chapter is to quantify four secondary metabolites from different *Ocimum* species: eugenol, ME, chavicol and estragole.

7.1.1 Chemical analysis of secondary plant metabolites

Plant secondary metabolites have evolved during plant development process to defend plants against bacteria, herbivores, viruses and fungi. Secondary metabolites serve as signal conductor to attract seed-dispersing animals and pollinating agents. They are immensely useful for humans to treat various health disorders and as pesticides, perfumes, spices as well (Wink, 2008). It is interesting to find out the single SM mode of action as plants are used as natural combinations in traditional medicine systems (Chevallier, 2001). Another interesting thing is to check which plant species have been used for medicinal purpose as some SM can be found in more than one plants, for example eugenol is present in cloves oil, basil, nutmeg and bay leaf (Khalil, *et al.*, 2017).

Secondary metabolites can be divided into three groups as shown in figure 7-1 (Croteau *et al.*, 2000). Around 200,000 secondary plant metabolite structures have been identified to date, a number that is growing continually (Hartmann, 2007). Several techniques such as HPLC (High-Performance Liquid Chromatography), TLC (Thin-Layer Chromatography), NMR (Nuclear Magnetic Resonance) and MS (Mass Spectrometry) are employed to separate and identify secondary metabolites (Lagrouh *et al.*, 2017).

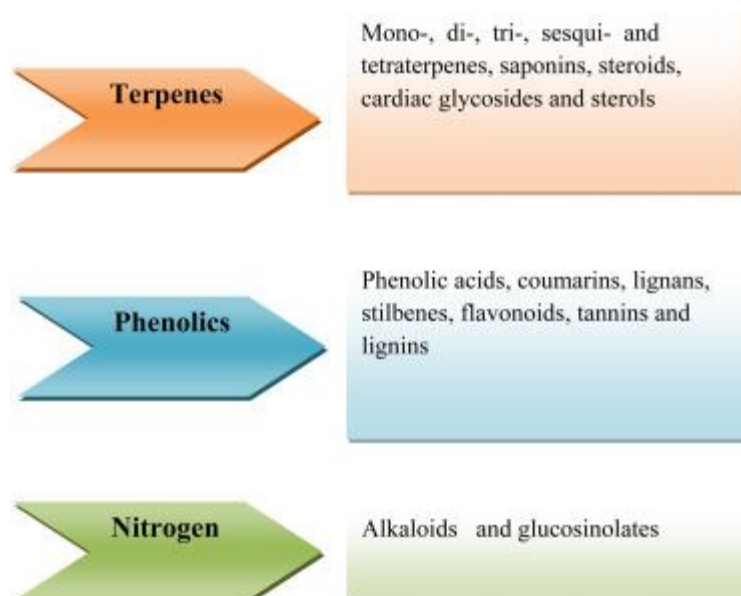


Figure 7-1: Plant phytochemical compounds (Ref.: <https://www.sciencedirect.com/science/article/pii/S1156523316303171>)

7.1.1.1 Chemical analysis in Tulsi plants

Essential oils are collectively known as volatile oils, as their compositions are different from normal oils and they are chemically unstable compounds. Unlike normal oils, essential oils are mainly obtained by steam distillation (Burt, 2004). Their presence can only be found in specially formed glandular glands or secretion cavities (Bakkali, 2008). Clove oil is the best example of such essential oil containing antiseptic properties that inhibit bacterial growth by interfering with their transportation and respiratory activities (Wei and Shibamoto, 2010). Essential oils contain oxygenated and hydrocarbon compounds and are used to store cosmetics and food items. Due to the presence of oxygenated compounds and aromatic rings, these essential oils contain specific odours. The aromatic compounds containing terpenes are completely soluble in alcohol and partially in water (Trease and Evans, 2009). An essential oil of plants mainly contains phenylpropenes and terpenoid compounds that are responsible for certain plant characteristics. Generally, an essential oil of plants (for example *Ocimum*) is directly used in phytotherapy, which contains a mixture of secondary metabolites. It is most likely that the use of these essential oils for medicinal purposes exhibits the synergetic interaction of several secondary metabolites present in the oil. Evaluation of a single compound is therefore difficult (Wink, 2015). Similarly, Tulsi essential oil also contains several secondary metabolites including eugenol, chavicol and their derivatives. These phenylpropenes have not been the subject of much study.

Along with eugenol, chavicol and their methyl derivatives, other phenolic compounds including cirsimaritin, rosmarinic acid, cirsilineol and apigenin are also present in various *Ocimum* species (Parasuraman *et al.*, 2015). These compounds are mainly found in essential oils from trichomes, while terpanoids, tannins, sterols, limelool, steroids and saponins are present in leaves and stems (Pattanayak *et al.*, 2010; Ali, 2016).

Two types of oil are present in the Tulsi: volatile oil, which has strong aroma and evaporates quickly, and non-volatile or fixed oil. Volatile and non-volatile oils occur mainly in leaves that are used in both fresh and dried forms (Jamshidi and Cohen, 2017). Eugenol, Methyl Eugenol (ME) and other phytoconstituents make up the majority of essential oils (Saharkhiz *et al.*, 2015). According to recent findings (Bhuvaneshwari *et al.*, 2016), eugenol and its derivatives make up around 45 per cent of the essential oils in *O. tenuiflorum*, while chavicol and methyl chavicol constitute approximately 25% of the essential oils. It has been reported that the production of these compounds reaches its maximum level in young leaves and during floral bud initiation, although their levels vary across *Ocimum* species (Saharkhiz *et al.*, 2015). In *O. basilicum*, the content of the compounds eugenol and ME were found to be correlated with plant heights. The amount of ME was predominant when plant heights were around 10cm, while taller plants contained high levels of eugenol (Miele *et al.*, 2001).

Before beginning chemical analysis of plant compounds, the material must be completely dried to ensure that the target compound is efficiently liberated into the solution. Air-, freeze- and oven-drying are the most popular techniques for this purpose. Freeze- and oven-drying techniques use high vacuum pressures and temperatures respectively, which are responsible for high losses of essential oil components, while air-drying naturally allows plant material to dry, consequently losing much less of the essential oils through evaporation because of the expansion of cells when freeze- or oven-drying, which allows oil to evaporate (Sasidharan, *et al.*, 2011). Studies have shown that eugenol levels remain stable during air-drying, with a very small decrease during freeze- and oven-drying (Díaz-Maroto *et al.*, 2004).

7.1.2 Chemical properties and threshold levels of target secondary metabolites

7.1.2.1 Eugenol (1-hydroxy-2-methoxy-4-allylbenzene)

Eugenol is a hydrophobic phenolic phenylpropane containing aldehyde and alcohol groups (Trease and Evans, 2009; Saharkhiz *et al.*, 2015). Eugenol plays an important part in Tulsi's medicinal properties, as it acts as an antioxidant, antihyperlipidemic and painkiller in dentistry (Pattanayak *et al.*, 2010; Singh *et al.*, 2012; Cohen, 2014). Eugenol is found in different amounts in each *Ocimum* cultivar (Jamshidi and Cohen, 2017). According to FAO (the UN's Food and Agriculture Organisation), a daily 2.5mg/kg eugenol intake is acceptable (Oetinger, 2003; Sağlam *et al.*, 2016; Yang, Zhao and Zeng, 2016).

The inhibition of glycation and advanced glycation (AGEs) end products are two of the main therapies used to cure diabetes. Eugenol has a dual mechanism which helps in this combination therapy. It also contains some anticancer properties: it can restrict melanoma cell proliferation, with a demonstrable 50% inhibition using 0.5 μ M. However, its isomer, including isoeugenol, does not show such activity. Gosh and his team (Ghosh *et al.*, 2005) carried out an experiment on laboratory rodents by using 125mg/kg eugenol, which decreased tumour size, delayed tumour growth and prevented tumour metastasis. It also exhibits an antibacterial effect (1mg/ml) against Gram+ bacteria (Laekeman *et al.*, 1990). A 2 μ /ml concentration of eugenol shows an inhibitory effect on *Helicobacter pylori* germs after nine to 12 hours of incubation, with a low pH value (Ali *et al.*, 2005). The development of bacterial resistance was not observed even after 10 passages, and this experiment also demonstrated prevention of *H. pylori* with sub-inhibitory eugenol concentration (Ali *et al.*, 2005). Devi *et al.* (2016) report that eugenol disrupts bacterial cell membranes and macromolecules present inside the cell.

Antifungal eugenol activity also been reported, where a eugenol dose of 10mg/kg/day was given to guinea pigs via an intravaginal route to treat *Microsporum gypseum*-type dermatophyte (Chami *et al.*, 2004a). A parallel experiment on male and female mice with liver tumours, carried out by administering eugenol at daily concentrations of more than 3,000ppm in their diets concluded that eugenol is a non-hepatocarcinog compound (National Toxicology Program 1983; Auerbach *et al.*, 2010). All these examples prove that eugenol acts as an active compound from a clinical point of view. Surprisingly, methyleugenol produced from eugenol is a known carcinogen and shows genotoxic activity, while eugenol's toxicity is considerably lower than its methyl derivative (Committee on H.M.P., 2011).

7.1.2.2 Methyl eugenol (ME)

Eugenol plays a precursor role in the formation of ME in the presence of the eugenol-O-methyltransferase enzyme (Bhuvaneshwari *et al.*, 2016). ME has similar characteristics to safrole, which is a carcinogenic phenylpropanoid (Telci *et al.*, 2006). ME's carcinogenicity has already been observed in animal models during clinical trials (Waddell, 2005). Activity related to ME's toxicity has also been observed in humans (Al-Subeihi *et al.*, 2015).

Basil, lemon grass, nutmeg, cloves and other culinary herbs also contain ME (Environment Canada, 2010), as do some fruits including bananas and grapefruit, at levels lower than 0.1mg/kg (TNO, 2010). The different amounts of ME levels in basil has been extensively studied (Lawrence *et al.*, 1988; Sheen *et al.*, 1991). Upper and lower ME dietary exposures have been decided by the Joint Committee of WHO and FAO in Europe and the US, these being 80.5µg/day and 9.6µg/day and respectively (JECFA, 2009; Williams and Mattia, 2009; IARC, 2013).

ME has been considered by the US Food and Drug Administration Act (21 CFR §172.515) as being less harmful (FDA, 2004), although in Europe and the UK the 1334/2008 regulations of 2011 have regulated threshold levels. The maximum level of ME for various food ingredients vary: 10mg/kg in sauces and soups, 15mg/kg in meat products, 1mg/kg in non-alcoholic drinks, 10mg/kg in fish products, 60mg/kg in readymade snacks and 20mg/kg in dairy products. Threshold levels of ME have not yet been decided for frozen and dried herbs. For example, large amounts of basil (which contains high levels of ME) are used in manufacturing pesto sauce (European Commission, 2008). Similar rules apply in Canada and the EU, where essential oils containing ME are used for cosmetic products. An exposure limit to ME is estimated at 0.01 per cent for fragrances, 0.0002 per cent for oral cleaning agents, 0.004 per cent for toiletries, 0.001 per cent for skin cleansers and 0.002 per cent for fragrance creams (SCCNFP, 2000; Health Canada, 2010). Canada's government has concluded that a maximum ME level of 200 µg/kg bw/day can be allowed for oral use (Government of Canada, 2010).

7.1.2.3 Chavicol

Chavicol, also known as *p-allylphenol*, is one type of organic compound. It contains the hydroxyl group that is constituted from the propenyl and hydroxyl groups. However, compared to other basil SMs, more study is still required to confirm the effects of this secondary metabolite.

7.1.2.4 Estragole (Methyl chavicol)

Estragole, also known as 1-allyl-4-methoxybenzene, belongs to the alkenylbenzen group and is a volatile phenylpropanoid. Table 7-1 shows the names of plants containing estragole. It is interesting that these plants also produce other metabolites that can affect estragole levels and characteristic in particular plant species.

Table 7-1: Amounts of estragole in plants essential oils (modified from EFSA, 2009, based principally on Council of Europe publications)

Sr. No.	Plant species	Estragole amount in %
1	<i>Anthriscus cerefolium</i>	Max 0.8
2	<i>Artemisia dranunculus</i>	0.7
3	<i>Foeniculum vulgar</i>	0.3
4	<i>Illicium verum</i>	Max 0.25
5	<i>Ocimum basilicum</i>	Approx. 0.4
6	<i>Pimpinella anisum</i>	Max 0.04

Estragole is a carcinogen (and therefore CVMOT expression is harmful) that has similarities to safrole. Though, it is only at dosages ranging from 1 to 10mg/kg bw, which is 100 to 1,000 times higher than human consumption, when laboratory rodents demonstrate estragole's genotoxic nature. It was therefore concluded that herbal products containing estragole do not pose a significant risk to human health at this stage. HMPC did, however, note that the use of products containing estragole should be restricted for certain groups of people such as pregnant and breastfeeding mothers and young children. HMPC also emphasises the necessity for further research on this component, as details regarding skin absorption are missing from the dataset (EMA, 2014).

According to HMPC guidelines, the average range of estragole levels in food products is 0.5-5mg, although this is different in each country. EU data, for example, indicates that daily exposure to estragole is as high as 1mg (CoE 2005), while in US it is around 0.6mg (JECFA 2009). In 2001, SCF concluded that estragole was a genotoxic and carcinogenic compound, so their recommendation was to reduce and restrict its usage. Later in 2002, the Flavor and Extract Manufacturers Association panel concluded that profiles of metabolic activity and covalent binding are only active at very high levels; at lower levels they are significantly less (Smith *et al.*, 2002). Since 2005, however, several research publications have mentioned estragole toxicity levels, prompting HMPC to reevaluate the toxicity levels of estragole. 0.5mg/person/day was determined for adults using herbal medicine treatments. This level has

only been accepted for the short-term (i.e. a maximum of 14 days) use of herbal medicinal products. In children, this level is set considerably lower – 0.2mg/day for body weights of up to 20kg.

7.1.3 Diode array detection (DAD) coupled to HPLC

A diode array is a detector coupled to an HPLC instrument which can detect the absorbance of light at several wavelengths simultaneously when a compound elutes from the HPLC column and passes through the flow cell (Swartz, 2010). The detector response should be proportionate to the compound concentration according to Beer-Lambert's Law, making it ideal for quantitative purposes (Mehta, 2012). As seen in figure 7-3, a DAD detector consists of one or several light sources generating a light that passes through the flow cell, where the eluting compounds from the HPLC column are introduced (Rance *et al.*, 2010). The grating (or a prism) will then disperse the light so that different wavelengths fall on different diodes.

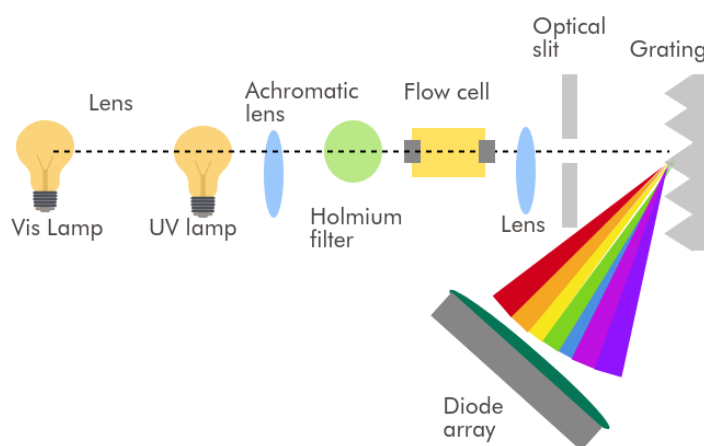


Figure 7-2: Schematic diagram of a diode array detector (DAD) (courtesy of Dr Linda Bengtstrom).

7.2 Materials and method

7.2.1 Chemicals and reagents

Acetonitrile (HPLC grade, 99.9 per cent) and methanol (HPLC grade, 99.9 per cent) was purchased from Rathburn (Walkerburn, UK). Mobile phase A was prepared using water (HPLC grade, 99.9 per cent) from Fischer Scientific (Loughborough, UK). Eugenol (>98 per cent), ME (>98 per cent) and methyl chavicol (>98 per cent) were purchased from Sigma-Aldrich

(Steinheim, Germany). Chavicol (>98 per cent) was obtained from abcr GmbH (Karlsruhe, Germany).

7.2.2 Samples analysed

Young leaves of *O. tenuiflorum* (both *Ram* and *Shyam* cultivars) and *O. gratissimum* were collected from De Montfort University's glasshouse. *O. basilicum* was purchased from a local vendor. All samples were air-dried at room temperature before plant extraction.

7.2.3 Sample preparation

After drying, the collected plant material (as per table 7-2) was manually ground using a mortar and pestle and were then left to macerate in 3mL methanol. The samples were then kept for 4°C O/N and were subsequently filtered using nylon filters (0.2µm) purchased from VWR (Lutterworth, UK). The samples were then transferred to clear HPLC glass vials for analysis.

Table 7-2: Sample preparation details

Sr. No.	Sample type	Sample weight	Solvent (MeOH)
1	Fresh	10 g	20 mL
2	Fresh	2 g	20 mL
3	Fresh	1 g	3 mL
4	Dry	1 g	5 mL
5	Dry	0.5 g	5 mL

7.2.4 Preparation of calibration standards

When working with HPLC, good calibration is absolutely essential to ensure reliable, quality results. Proper calibration of an HPLC instrument begins with the making of a suitable calibration standard. Here, calibration standards were prepared for an eight-point calibration curve according to Table 7-3.

Table 7-3: Preparation of different stock solution concentrations to obtain a calibration curve for each standard

Compound	STD 1 (µg/ml)	STD 2 (µg/ml)	STD 3 (µg/ml)	STD 4 (µg/ml)	STD 5 (µg/ml)	STD 6 (µg/ml)	STD 7 (µg/ml)	STD 8 (µg/ml)
Eugenol	5	2.5	0.625	0.16	0.08	0.02	0.00	0.00
ME	50	25	6.25	1.56	0.78	0.20	0.05	0.00
Chavicol	1000	500	125	31.25	15.63	3.91	0.98	0.00
Methyl chavicol	1000	500	125	31.25	15.63	3.91	0.98	0.00

7.2.5 HPLC-DAD analysis

Analyses of the secondary metabolites were performed using an Agilent 1100 series HPLC with a DAD (Santa Clara, CA, US). The column used was Kinetex® 5C18 (5µm, 4.6 x 150 mm) with a KrudKatcher Classic HPLC In-Line Filter 0.5µm filter (Phenomenex, Torrance, CA, USA). Mobile phase ‘A’ was water and 0.1% acetic acid, and mobile phase ‘B’ was acetonitrile and 0.1 per cent acetic acid. This mobile phase mixture was used for an isocratic separation in 50:50 proportions at a total run time of 15 minutes. The column temperature was set to 35°C and flow rate was 1mL min⁻¹. Injection volume was 5µL and the detector wavelength was set to 250nm.

7.3 Results

7.3.1 Calibration and validation of HPLC method

A total of four secondary metabolites (eugenol, ME, chavicol and methyl chavicol) were extracted from *O. basilicum*, *O. tenuiflorum* Shyam, *O. gratissimum* and *O. tenuiflorum* Ram leaves and were analysed by HPLC-DAD. Figure 7-4 (A-D) illustrates typical calibration curves obtained for eugenol, ME, chavicol and methyl chavicol. The linear regression (R^2) at 0.99 for all analytes indicates a satisfactory linearity for all compounds analysed.

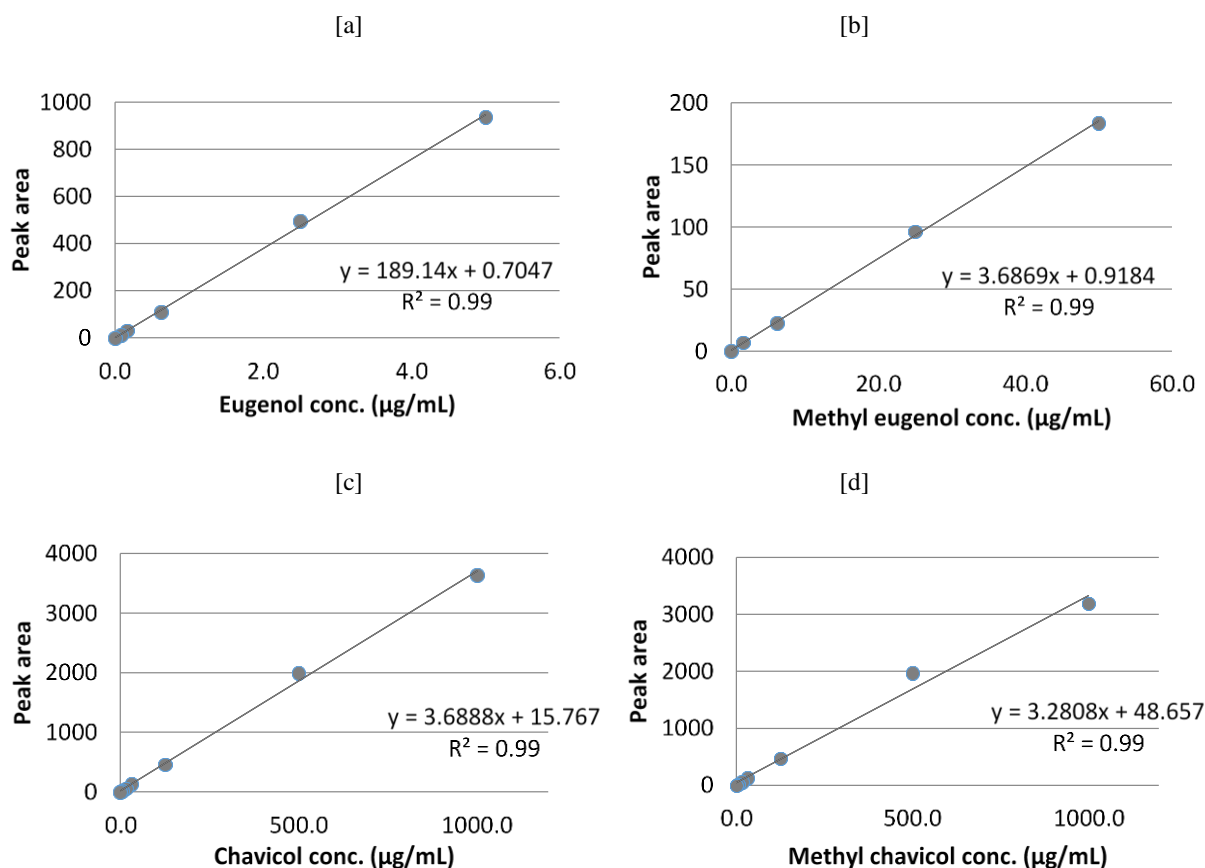


Figure 7-3: Examples of calibration graphs obtained for [a] eugenol [b] ME [c] chavicol and [d] methyl chavicol concentration vs HPLC peak area

The limits of detection (LOD), defined as three times the standard deviation for the lowest standards in the calibration curve, were established as follows: eugenol was 22 ng mL^{-1} , ME $1\text{ }\mu\text{g mL}^{-1}$, chavicol $7\text{ }\mu\text{g mL}^{-1}$ and methyl chavicol $1\text{ }\mu\text{g mL}^{-1}$. The limit of quantification (LOQ), defined as ten times the standard deviation for the lowest standards in the calibration curve, were established as follows: eugenol was 75 ng mL^{-1} , ME $3.5\text{ }\mu\text{g mL}^{-1}$, chavicol $23\text{ }\mu\text{g mL}^{-1}$ and methyl chavicol $3.5\text{ }\mu\text{g mL}^{-1}$.

The representative chromatogram shown in figure 7-5 illustrates the separation of the four secondary metabolites investigated in *O. tenuiflorum* Ram. As seen in Figure 7-5, the smallest and most polar compounds, chavicol followed by eugenol are eluted first in the reverse phase HPLC separation. Both of these compounds contain an alcohol (-OH) as a functional group, which can increase the polarity of the molecule (Koeduka *et al.*, 2006). The interactions between the analyte and stationary phases increase along with the non-polarity increases, as is the case with ME and methyl chavicol, leading to a longer retention time for these compounds

(Tshilanda *et al.*, 2016). This increase in non-polarity is due to the addition of a methyl group (-OCH₃) to these compounds. Moreover, no co-eluting components were detected during the evaluation of the specificity of the method. It should also be noted that all analytes are baseline-separated.

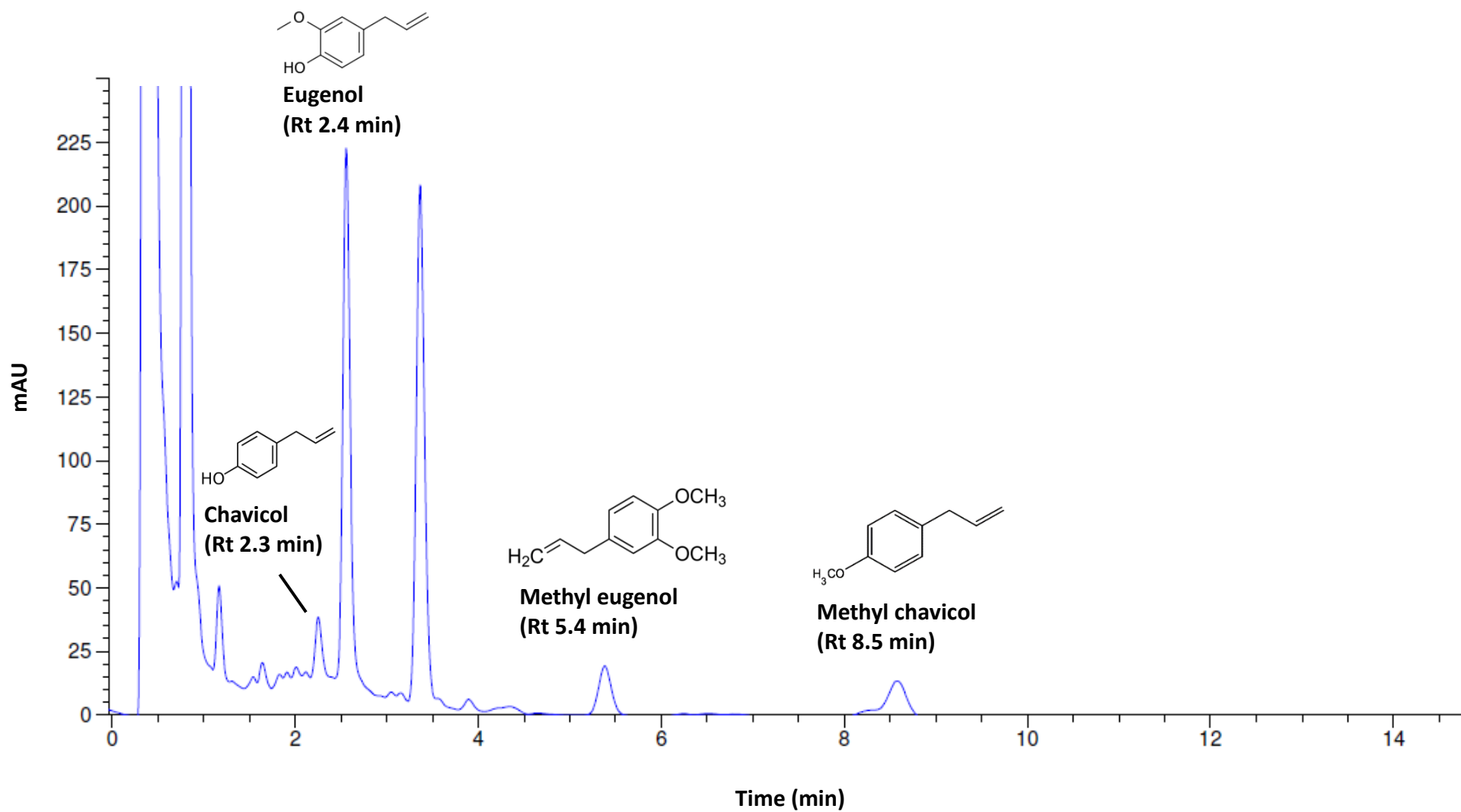


Figure 7-4: A representative chromatogram of eugenol, ME, chavicol and methyl chavicol extracted from dried *O. tenuiflorum* Ram leaves and extracted by methanol

7.3.2 Optimisation of extraction

Eugenol is an established antidiabetic agent, and an aqueous extract of Tulsi leaves has been reported to be used by the local population in India (Naga Raju *et al.*, 2006). As seen in figure 7-5, all the secondary metabolites analysed contain polar functional groups. According to the principle “like dissolves like” a polar solvent such as methanol is suitable for the extraction of eugenol, ME, chavicol and methyl chavicol from plant material.

Initial pilot data (see figure 7-6) reveals that the most efficient method for extraction of the analytes from *O. basilicum* is to use dry leaves rather than fresh. In general, a smaller amount of material, either fresh or dry, gave higher extraction efficiency for the secondary metabolites analysed. These findings are in accordance with results published by Díaz-Maroto *et al.* (2004), which show that air-drying *Ocimum* leaves does not affect eugenol concentration (ranged from 40-60 $\mu\text{g/g}$).

Even though no concentrations of methyl chavicol were found in any of the *O. basilicum* extracts, it can be assumed that extraction efficiency will be akin to the chemically similar secondary metabolites found in the extract.

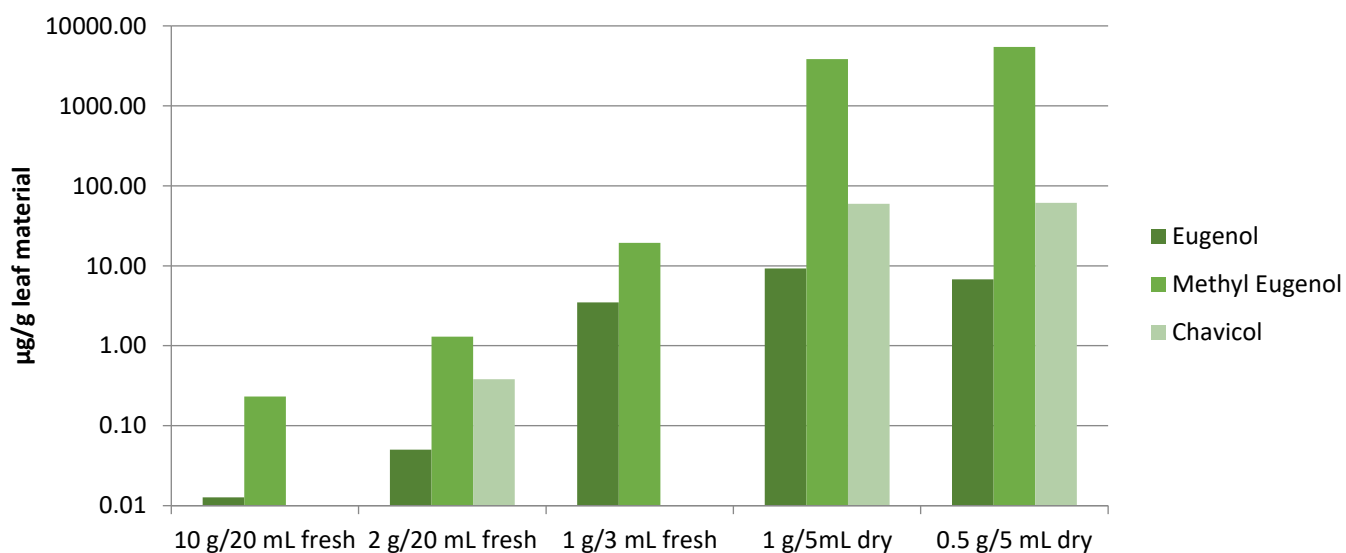


Figure 7-5. Concentrations ($\mu\text{g g}^{-1}$) of eugenol, ME and chavicol, measured in *O. basilicum* fresh and dry leaves in different proportions to the solvent volume. Note that the y-axis is a logarithmic scale

Both fresh and dried samples are used in this study to determine whether the drying process affect the SM levels. The comparison between fresh and dry leaves along with different solvent volume showed more SM production in dry samples.

Dried plant materials are common and best to extract phytochemicals as dry plant material gives free bonding of solvents with the phytochemicals according to its polarity. Wet sample produce more viscosity and lack separation of secondary metabolites corresponding to polarity of the solvent used.

7.3.3 Analysis of Tulsi extracts

Once the optimal extraction conditions were established (see Section 7.3.2 using *O. basilicum* as a model system), samples collected from *O. basilicum*, *O. gratissimum*, *O. tenuiflorum* Ram and *O. tenuiflorum* Shyam were extracted according to the optimised protocol. As seen in Figure 7-6, the highest concentration of eugenol was found in *O. tenuiflorum* Ram ($21\mu\text{g g}^{-1}$). In a study by Annad *et al.* (2016), the highest amounts of ME were reported in *O. tenuiflorum*, while this experimental dataset had shown the most abundant ME in *O. basilicum* which was around $1,800\mu\text{g}$ in each gram of plant material. As seen in figure 7-7, the results obtained in terms of the proportions of eugenol and ME between these two species obtained in this study were in accordance with these previously published results. Moreover, since a smaller amount of eugenol was detected in *O. basilicum* plant extracts than in *O. gratissimum* and *O. tenuiflorum* Ram, it is possible that a majority of eugenol had been converted to ME.

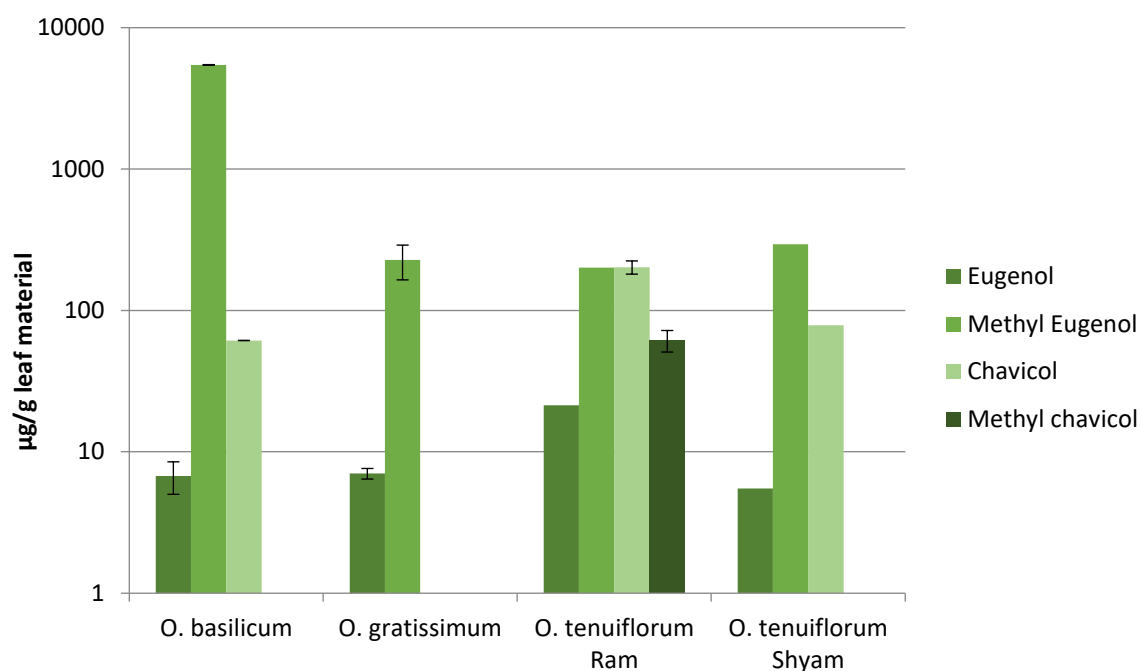


Figure 7-6: Concentrations ($\mu\text{g g}^{-1}$) of eugenol, ME, chavicol and methyl chavicol measured in dried leaves from *O. basilicum*, *O. gratissimum*, *O. tenuiflorum* Ram and *O. tenuiflorum* Shyam. Standard deviation is indicated as error bars. Note that the y-axis is a logarithmic scale

The highest concentration of ME was found in *O. tenuiflorum* Ram ($202 \mu\text{g g}^{-1}$) (Figure 94). However, no chavicol was detected in *O. gratissimum*. These findings are in conflict with previously reported results, where chavicol was detected in *O. gratissimum* and the highest amount was detected in *O. basilicum* (Anand *et al.*, 2016).

Additionally, $60\mu\text{g}$ of methyl chavicol was detected in 1g of *O. tenuiflorum* Ram leaf material (Figure 7-6). No concentrations of methyl chavicol were detected in any of the other plant species. This finding is also in conflict with previous studies, which found methyl chavicol only in *O. basilicum* (Anand *et al.*, 2016).

7.4 Discussion

The presence of eugenol and chavicol and their methyl derivatives in basil is likely to contribute to the antimicrobial, antidiabetic and anti-carcinogenic properties of the *Ocimum* plant. Eugenol is the most-studied antidiabetic agent of all the secondary metabolites, and therefore varieties containing eugenol can act as a remedy for diabetic patients. Plant material, mainly

leaves, from the *Ocimum* genus can be used in fresh or dry form to produce these herbal remedies.

Ocimum is an important genus of the Lamiaceae family. It consists of aromatic annual and perennial herbs and bushes (Boggia *et al.*, 2014). Different *Ocimum* species will produce varying concentrations of phenylpropanoids, such as eugenol, chavicol and their methyl derivatives. In this study, all four secondary metabolites were found in *O. tenuiflorum* Ram. Eugenol and ME were found in leaves from the *O. basilicum* and *O. tenuiflorum* Shyam plants.

Additionally, the chavicol and methyl chavicol are both absent in *O. gratissimum*, which might indicate the absence of *CVOMT* enzyme activity and also of substrate. Conversely, in *O. tenuiflorum* Shyam and *O. basilicum*, the substrate for methyl chavicol, chavicol were found, yet the absence of the relevant enzyme (*CVOMT*) might affect the production of methyl chavicol. All these compounds are synthesised on the glandular trichomes of the leaves of plants in the *Ocimum* genus, a synthesis that is related either to the structure of these leaves (Gang *et al.*, 2001) or to plant heights (Miele *et al.*, 2001).

The diversity of the various phenylpropanoids in the *Ocimum* species can be explained by the differential gene expression patterns of *EOMT* and *CVOMT*. The low expression levels of particular enzymes might lead to the absence of related compounds. It might also have some connection with the post-translation modifications in which, certain metabolites are absent because of transcript abundance. For example, the amount of chavicol is higher in *O. basilicum*, yet due to post-translation ubiquitylation of the *CVOMT* enzyme, no methyl chavicol was detected (Xie *et al.*, 2008).

7.5 Conclusion

In general, all the *Ocimum* plants studied showed some level of production of the antioxidants eugenol, chavicol and their methyl derivatives, which is beneficial for treating the number of diseases. However, plant extracts and their isolated compounds cannot be similar to Tulsi used as a whole plant in its natural form. Tulsi has various therapeutic benefits apart from those relating to diabetes. It has been included in diets, too. A reconsideration of its use alongside synthetic drugs is, however, recommended.

7.5.1 Further work

In this study, a target chemical analysis of the four secondary metabolites (eugenol, ME, chavicol and methyl chavicol) extracted from dried leaves from various plants of the *Ocimum* genus and analysed by HPLC-DAD has been presented. Future studies could analyse these phytochemicals in other *Ocimum* species. Moreover, plant material could be collected from different parts of the plant and at different stages in its development in order to further investigate the production of the phytochemicals. Another study might chemically analyse commercial products including powdered Tulsi or Tulsi tea to further establish the health benefits of such products. These antioxidant compounds present in *Ocimum* plants can also be further analysed by designing lengthier DPPH assays. A test for the adverse health effects of the *Ocimum* plant extracts and commercial products produced using various human cell assays by designing lengthier DPPH assays is also recommended.

8 General Discussion and future work

8.1 Summary of the research project

This research project was set out to compare the AA sequences of the enzyme *O-methyltransferase* within different *Ocimum* species. The results of this study have produced a novel approach by identifying lesser-known secondary methods of metabolite production and how they differ within each species and chemotypes. Furthermore, molecular analysis including DNA and AA sequences were studied in great detail in order to provide a firm foundation for sequence analysis and the study of protein production (Chapter 4).

Authentication, sequence analysis and different phases of the research study are illustrated in Figure 8-1.

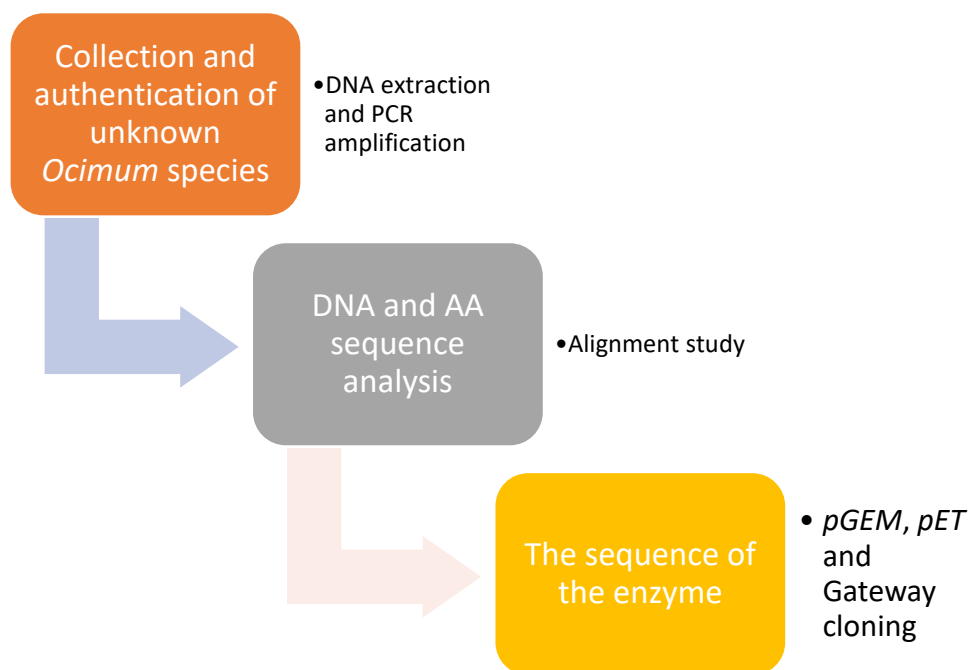


Figure 8-1: Different phases of the research study. The figure highlights the direction of the entire research study, from the point of identifying the *OMT* gene in the *Ocimum* species as well as designing different cloning strategies including G-blocks for the target *Ocimum* species

Tulsi has great value within the Hindu community as a sacred plant. It is famous for its numerous medicinal benefits and its religious significance, although this research study has targeted different properties of this plant by identifying some lesser-known enzymes which expression can be harmful to humans. The plant's commercial value can be increased by removing or minimising the effects of these genes in certain *Ocimum* species.

8.1.1 Authentication of the starting material

The plastid *trnH-psbA* region has been selected for authentication purposes due to its high copy number, good priming sites, interspecific variation and, most importantly, the short region which may be useful for degraded and damaged samples (Kress *et al.*, 2005; Howard, 2010). The DNA-based identification method has begun to gain popularity over the established chemical identification method. Three types of Tulsi were initially collected from Hindu community: *Ram* (*O. tenuiflorum*), *Shyam* (*O. tenuiflorum*) and *Vrinda* (*O. gratissimum*). The main aim of this research was firstly to identify unknown samples using the DNA barcoding technique. Once this technique was in place, *OMT* genes were isolated using the same strategies from fresh plants, vouchered specimens and commercial samples. Previous studies have revealed that the *trnH-psbA* plastid region is more reliable for identification of *Ocimum* species than the ITS region (Bhamara *et al.*, 2015), a finding further supported by the present research study's identification of *OMT* genes in authenticated *Ocimum* samples.

8.1.2 Explore the occurrence of *OMT* DNA polymorphism

After authentication, targeted *OMT* primers were first identified and tested to isolate the *OMT* gene from a range of *Ocimum* species. Renu *et al.* (2014) and Gang *et al.* (2001 and 2002) findings were used to identify suitable *OMT* primer pairs. As well as these literatures, some primers were also designed using online primer design tools. Among all these primer pairs, the *EOMT* primer sequence taken from Renu *et al.* (2014) proved most effective and was used throughout the study. This *EOMT* primer pair has managed to cover most of gene region, although it could not efficiently cover the middle, the beginning and end parts of the sequence. By designing two more primer pairs, entire *OMT* genes (1074 bp) were amplified. These interesting findings led this research to obtain novel results which have never been achieved before. This was done by using these primer pairs for sequence purpose. A detailed alignment containing different *Ocimum* species and chemotypes was prepared in order to study DNA polymorphism of the *OMT* gene. It is the first time this kind of alignment containing the different *Ocimum* species and chemotype sequencing has been prepared. As different chemotypes are included for a single species (*O. tenuiflorum*), it is easier to see the differences and similarities between the base pairs. DNA sequences were subsequently translated into AA sequences. After carefully studying of the AA alignment to determine

their differences, the relevant DNA sequences were further cross-checked. After preparing the final alignment output, polymorphism tables were prepared to identify AA differences within different *Ocimum* species and varieties of *O. tenuiflorum*. Some visible differences can be seen in the form of different groups. *O. tenuiflorum* and *O. gratissimum* display some similarities by their location in one group, while other species including *O. basilicum*, *O. americanum*, *O. citriodorum*, *O. kilimandscharicum* and *O. selloi* can be placed in another group by examining the AA differences within these groups. Gang *et al.* (2002) shows only 261 positional differences (Figure 8-2) responsible for *EOMT* and *CVOMT* gene structures. This study, on the other hand, reveals many other combinations that can be further studied to find AA polymorphisms within different *Ocimum* species. This unique finding can provide a strong foundation for much further research, as this kind of research on *OMT* gene has never before been conducted on such a large scale. By checking AA arrangements, protein production levels were further studied in the bacterial system. The resulting findings can be used for plant transformation studies, either by using *Arabidopsis* or *Ocimum* plant systems where *OMT* genes might be modified to obtain more benefits from this valuable medicinal plant.

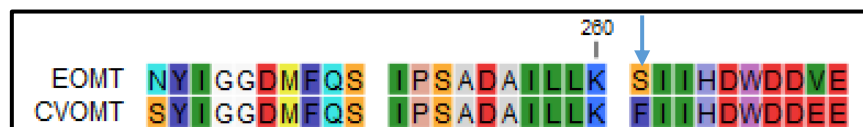


Figure 8-2: Ser-261 represents *EOMT*, while Phe-261 represents *CVOMT* in Sweet Basil (*O. basilicum*) (Gang *et al.*, 2002)

AA No.	20	42	51	62	68	72	73	74	79	90	93	98	134	139	140	143	155	163	167	178
<i>EOMT (O. basi)</i>	V	G	G	S	E	C	F	Q	A	N	N	C	H	F	T	K	T	E	G	S
<i>CVOMT (O. basi)</i>	A	G	D	A	E	S	F	Q	A	N	N	C	Y	F	K	N	T	K	G	S
<i>O. basi</i> (B16)	V	G	D	A	E	S	F	Q	A	N	N	C	Y	F	K	N	T	K	G	S
<i>O. basi</i> (G34)	A	G	D	S	E	S	F	Q	A	N	N	C	Y	F	K	N	T	K	G	S
<i>O. citri</i> (G29)	A	G	D	A	E	S	F	Q	A	-	N	C	Y	F	K	N	T	K	G	S
<i>O. ameri</i> (C85)	A	G	D	A	E	S	F	Q	A	-	N	C	Y	F	K	N	T	K	G	S
<i>O. klima</i> (G31)	A	D	D	A	E	S	F	Q	A	-	-	C	Y	F	K	N	T	K	G	S
<i>O. selloi</i> (C82)	A	G	G	S	E	C	F	H	A	N	N	C	H	F	K	K	T	E	N	S
<i>O. grati</i> (G30b)	A	G	G	S	A	C	L	H	V	N	N	C	H	L	T	H	M	E	S	S
<i>O. grati</i> (G32)	A	G	G	S	A	C	L	H	V	N	N	R	H	L	T	N	T	E	G	S
<i>O. grati</i> (C81)	A	G	G	S	A	C	L	H	V	N	N	R	H	L	T	H	M	E	S	S
<i>O. tenu</i> (G37)	A	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (14092902)	A	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (14092909)	A	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (G26)	A	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (G27)	A	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (G36)	A	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (G38)	A	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (G39)	A	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (V86)	A	G	G	S	E	C	L	H	A	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (14092903-ME)	-	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	T	E	G	A
<i>O. tenu</i> (V87-ME)	A	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (V88-E)	A	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (V92)	A	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (V93)	A	G	G	S	A	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (V94)	A	G	G	S	A	C	L	H	V	K	N	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (V98)	A	G	G	S	E	C	L	H	V	K	S	H	H	F	T	H	M	E	G	A
<i>O. tenu</i> (V99)	A	G	G	S	A	C	L	H	V	K	N	H	H	F	T	H	M	E	G	A

Continue on next page...

AA No.	194	206	217	220	231	239	249	261	264	265	267	270	271	284	299	326	327	330	340
<i>EOMT</i> (<i>O. basi</i>)	E	N	M		V	N	Q	S	H	D	D		G	M	H	M	S	E	S
<i>CVOMT</i> (<i>O. basi</i>)	D	N	V	M	V	K	Q	F	H	D	D		G	I	H	M	N	E	S
<i>O. basi</i> (B16)	D	N	V	M	V	K	Q	S	H	D	D		G	I	H	M	N	G	S
<i>O. basi</i> (G34)	D	N	V	M	V	R	Q	S	H	D	D		G	I	H	M	N	G	S
<i>O. citri</i> (G29)	D	N	M	M	V	K	Q	S	H	D	D		G	V	H	M	N	G	S
<i>O. ameri</i> (C85)	D	N	V	M	V	K	Q	F	H	D	D		G	I	H	M	N	E	S
<i>O. klima</i> (G31)	D	N	V	M	V	R	Q	S	H	D	D		G	I	H	M	N	G	S
<i>O. selloi</i> (C82)	E	D	V		V	R	Q	C	H	D	N		S	T	H		K	A	S
<i>O. grati</i> (G30b)	Q	N	V	M	V	T	Q	S	H	D	N		G	I	Q		K	A	S
<i>O. grati</i> (G32)	m	N	V	M	V	S	R	S	H	D	N		G	I	H		K	A	S
<i>O. grati</i> (C81)	m	N	V	M	V	S	R	S	H	D	N		G	I	H		K	A	S
<i>O. tenu</i> (G37)	G	D	V		V	S	R	S	H	D	N		G	I	Q		K	A	G
<i>O. tenu</i> (14092902)	G	D	V		V	S	R	S	H	D	N		G	I	Q		K	A	G
<i>O. tenu</i> (14092909)	G	D	V		V	S	R	S	H	D	N		G	I	Q		K	A	G
<i>O. tenu</i> (G26)	E	D	V	M	V	S	R	S	H	D	N		G	I	Q		K	A	G
<i>O. tenu</i> (G27)	G	D	V		V	S	Q	S	Q	D	N		G	M	Q		K	A	G
<i>O. tenu</i> (G36)	E	N	V		V	S	R	S	Q	D	N		G	M	Q		K	A	G
<i>O. tenu</i> (G38)	G	D	V		V	S	R	S	H	D	N		G	I	Q		K	A	G
<i>O. tenu</i> (G39)	G	D	V		V	S	R	S	H	D	N		G	I	Q		K	A	G
<i>O. tenu</i> (V86)	E	D	V		M	S	R	S	H	D	N		G	I	Q		K	A	G
<i>O. tenu</i> (14092903-ME)	G	D	V		V	S	R	S	H	D	N		G	I	Q		K	A	G
<i>O. tenu</i> (V87-ME)	G	D	V		V	S	R	S	H	D	N		G	I	Q		K	A	G
<i>O. tenu</i> (V88-E)	G	D	V		V	S	R	S	Q	D	N		G	I	Q		K	A	G
<i>O. tenu</i> (V92)	G	D	V		V	S	R	S	H	D	N		G	I	Q		K	A	G
<i>O. tenu</i> (V93)	E	D	V	M	M	S	R	S	H	Y	N		S	I	Q		K	A	G
<i>O. tenu</i> (V94)	E	D	V		V	S	R	S	H	D	N		G	I	Q		K	A	A
<i>O. tenu</i> (V98)	G	D	V	M	V	S	R	S	Q	D	N		G	I	Q		K	A	G
<i>O. tenu</i> (V99)	E	D	V		V	S	R	S	H	Y	N		S	I	Q		K	A	A

Figure 8-3: AA alignment outputs showing their arrangements in different *Ocimum* species within two outputs, from AA No. 1 to 340. AA similarities and differences are highlighted in this figure. The red highlighted boxes show the similarities between *O. tenuiflorum* and *O. gratissimum* sample sequences at AA Positions 51, 62, 72, 73, 74, 163, 217, 261, 267, 326, 327 and 330. The combination of red and black boxes shows the differences between *O. tenuiflorum* and *O. gratissimum* at AA Positions 90, 93 and 139. The black boxes at AA Positions 51, 72, 73, 74, 98, 134, 140, 143, 155, 163, 178, 194, 220, 249, 267, 299, 326, 327 and 340 show the similarities and differences within *O. basilicum*, *O. americanum*, *O. citriodorum*, *O. kilimandscharicum* and *O. selloi*.

AA No.	5	6	20	51	61	62	68	72	73	74
EOMT (O. basi)	K	V	V	G	Q	S	E	C	F	Q
CVOMT (O. basi)	N	M	A	D	K	A	E	S	F	Q
O. grati (1138)	K	V	A	G	Q	S	A	C	L	H
O. tenu (285/866)	K	V	A	G	Q	S	A	C	L	H

AA No.	79	90	91	93	98	106	111	117	134	139
EOMT (O. basi)	A	N	N	N	C	C	E	I	H	F
CVOMT (O. basi)	A	N	-	N	C	R	G	A	Y	F
O. grati (1138)	V	N	N	N	R	R	G	A	H	L
O. tenu (285/866)	V	K	N	S	H	R	G	A	H	R

AA No.	140	143	155	163	166	167	178	194	206	217
EOMT (O. basi)	T	K	I	E	K	G	S	E	N	M
CVOMT (O. basi)	K	N	T	K	M	G	S	D	N	V
O. grati (1138)	T	H	M	E	M	S	S	E	N	V
O. tenu (285/866)	T	H	M	E	M	G	A	G	D	V

AA No.	220	239	241	249	261	267	269	283	284	299
EOMT (O. basi)	I	N	N	Q	S	D	V	V	M	H
CVOMT (O. basi)	M	K	S	Q	P	D	E	G	I	H
O. grati (1138)	M	S	S	R	S	S	E	G	I	H
O. tenu (285/866)	I	S	S	P	S	N	E	G	I	Q

AA No.	317	326	327	330	334	335	339	340
EOMT (O. basi)	C	M	S	E	Y	D	K	G
CVOMT (O. basi)	S	M	N	E	S	A	T	G
O. grati (1138)	C	E	K	A	S	D	T	S
O. tenu (285/866)	C	E	K	A	S	D	T	G

Figure 8-4: Polymorphic tables showing several AA similarities and differences within four sequences of *Ocimum* species used to prepare G-blocks. The *O. gratissimum* and *O. tenuiflorum* EOMTs match the *O. basilicum* EOMT (but not the CVOMT) at AA Positions 5, 6, 20, 51, 61, 72, 134, 140, 163, 261, 317 and 335. At AA Positions 106, 111, 117, 166, 217, 241, 269, 283, 284, 334 and 339 the sequences of *O. tenuiflorum* and *O. gratissimum* match the *O. basilicum* CVOMT but not the *O. basilicum* EOMT.

By preparing an AA alignment using the sequences of different *Ocimum* species, the OMT similarities and differences in the gene sequence were studied in detail. The sample orders were arranged based on phylogenetic analysis, where *O. basilicum* *EOMT* and *O. basilicum* *CVOMT* reference sequences were arranged on top followed by *O. basilicum*, *O. citriodorum*, *O. americanum*, *O. kilimandscharicum* and *O. selloi* sequences. All *O. gratissimum* and *O. tenuiflorum* sequences were later collated. As described earlier, AA Position 261 was highlighted in previous research (Gang *et al.*, 2002) between *EOMT* and *CVOMT* gene sequences in *O. basilicum* only. The current research findings take the next step where several combinations of AA arrangements were found within seven *Ocimum* species. As shown in Figure 8-3, the similarities were highlighted within *O. gratissimum* and *O. tenuiflorum* sequences at several AA positions. At the same positions, the AA were found to be different from them within the remaining five *Ocimum* species, including two published *O. basilicum* sequences. There are some places where *O. gratissimum* and *O. tenuiflorum* AA display differentiation. Similarly, *O. basilicum*, *O. americanum*, *O. citriodorum*, *O. kilimandscharicum* and *O. selloi* AA sequences showed interesting combinations at the highlighted places. The general conclusion can be derived from these findings that *O. gratissimum* and *O. tenuiflorum* sequences evolved at the same evolutionary stage, and therefore show striking similarities within their AA sequence arrangements. The remaining *Ocimum* sequences can be related to each other, though they have surely evolved distantly from *O. gratissimum* and *O. tenuiflorum*.

The second most important finding of this research regards the identification of an *EOMT* gene in different *Ocimum* species. The *EOMT* gene sequences from *O. tenuiflorum* and *O. basilicum* are already present in the databases, although this is the first time an *EOMT* gene sequence has been identified from *O. gratissimum*, *O. americanum*, *O. citriodorum*, *O. kilimandscharicum* and *O. selloi* species.

8.1.3 *OMT* gene transformation in *E. coli*

In *Ocimum*, *OMT* is one of the key genes involved in the methylation process, especially in young leaves (Renu *et al.*, 2014). In the present work the *OMT* gene was tested by bacterial transformation study and chemical analysis. Previous studies have shown that certain secondary metabolite production levels vary in different *Ocimum* species (Anand *et al.*, 2016). This finding led to another study of the reasons why *OMT* protein production levels

vary across *Ocimum* species. It may be due to the inactivity of enzyme caused by that decrease, or sometimes to the cessation of production of methyl derivatives such as methyl eugenol and estragole (Anand *et al.*, 2016). Based on this finding, the present bacterial transformation study has been designed to examine *OMT* protein levels in four *Ocimum* species: *O. basilicum* (*EOMT*), *O. basilicum* (*CVOMT*), *O. tenuiflorum* and *O. gratissimum*.

The *pGEM* vector transformation was designed by using a poly A tail approach. Molecule stability can be increased during RNA processing by adding a poly tail containing multiple adenine residues to mRNA. This also prevents the degradation of RNA and facilitates the process by which mature RNA is exported from the nucleus and translated into protein. For screening the recombinant clones, blue-white colony screening was used for transformed colonies using the *pGEM* vector. For the *pET* vector, a strategy of restriction enzymes was used, that involved identifying two REs at the 5' and 3' ends. By using double digest reaction, GOIs were inserted in the correct orientation into the recipient plasmid. The third cloning strategy used in this study was Gateway cloning, which differs from the others. In Gateway cloning, specially designed 'Gateway *att*' sites and enzyme mixes were used. During the cloning process the GOI was transferred from one vector to another while keeping the reading frame in the correct orientation. For *pET* vector transformation and the Gateway cloning strategy, the antibiotic resistance (kanamycin) technique was used to screen colonies. This is unusual that three cloning strategies have been used for one research study, and the results are impressive. All three cloning result outputs are clear and remarkable, confirming that the project is proceeding in the right direction. By using these results and the products prepared at the end of each cloning, the next stage of the study of protein expression and plant transformation can be initiated without repeating any of the cloning procedures.

The *pGEM* vector was used to obtain several transformed colonies, the *pET* vector was used to study protein production, and the Gateway cloning system was used to prepare the binary vectors containing GOIs for the further study of plant transformation (Figure 8-5).

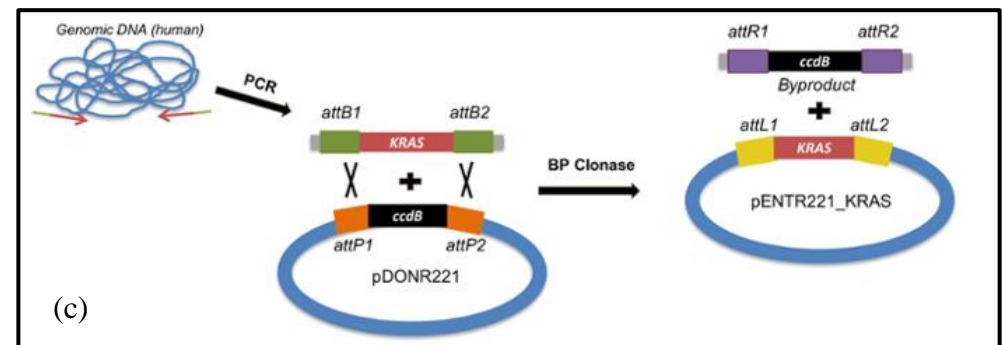
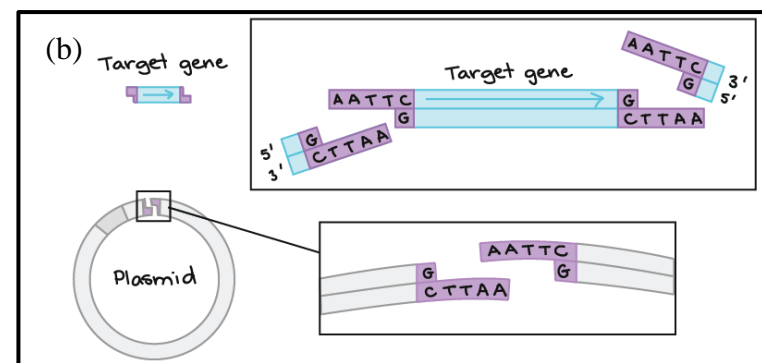
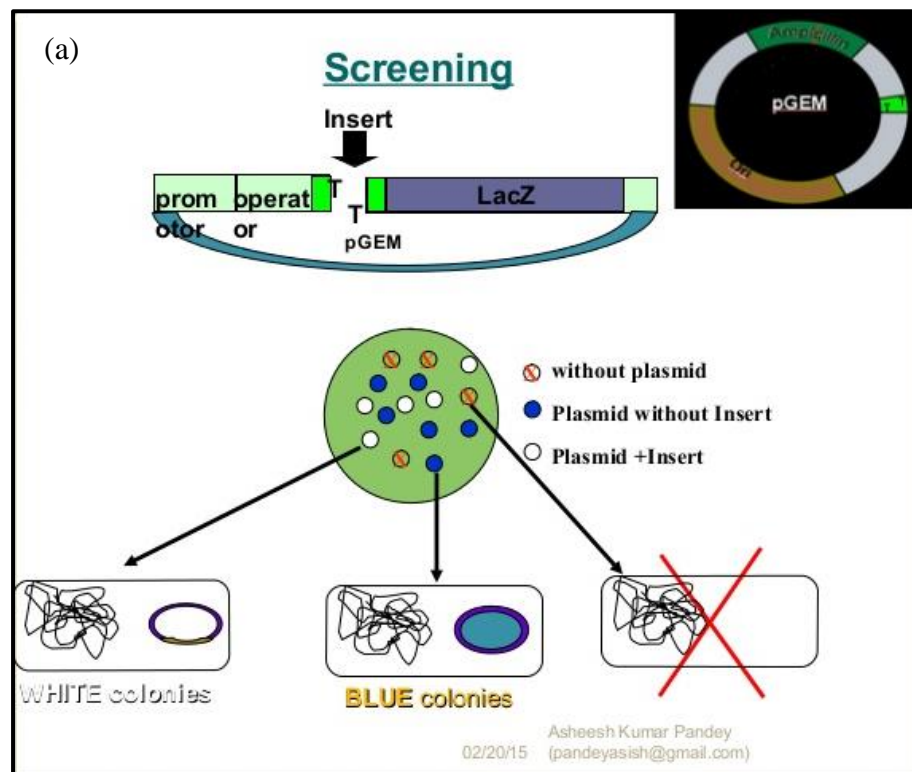


Figure 8-5: Different cloning strategies. (a) pGEM (Ref.: <https://www.slideshare.net/guddasish2007/gene-cloning-44913843>; (b) pET (Ref.: <https://www.khanacademy.org/science/biology/biotech-dna-technology/dna-cloning-tutorial/a/restriction-enzymes-dna-ligase>); (c) Gateway cloning strategies (Ref.: <https://blog.addgene.org/plasmids-101-gateway-cloning>).

8.1.4 The identification of secondary metabolites through chemical analysis

Ocimum, a member of the Lamiaceae family, contains a diversity of phenylpropanoids, of which *O. gratissimum* and *O. tenuiflorum* contain large amounts of eugenol and its methyl derivatives, while estragole is the most abundant in *O. basilicum* (Raina *et al.*, 2013). The aerial parts of the plant contain the glandular trichomes where these major phenylpropanoid compounds are synthesised. They are found mainly in young leaves and inflorescence, and their chief role is to protect the plant against pathogen attack (Vassão *et al.*, 2006). They are less prevalent in mature plant tissues, and only a negligible amount of these compounds is found in the roots and rhizomes (Wink, 2010). According to Anand *et al.* (2016), different gene expression patterns are responsible for metabolite diversity within various *Ocimum* species. Based on this literature study, an *OMT* gene has been studied with regard to high and low levels of metabolite production. As well as considering the levels of secondary metabolites, differences in the protein sequence that could affect enzyme activity was studied in detail. For example, the various gene activities of *OMT* enzymes in *O. basilicum* appear to control the production of the relevant phenylpropanoids (Xie *et al.*, 2008). Apart from these factors, enzyme levels during the transcription process are also responsible for the variation in metabolite accumulation (Zvi *et al.*, 2012; Singh *et al.*, 2015). For example, in chemotypes of *O. basilicum*, different levels of phenylpropanoid have been observed by previous studies (Xie *et al.*, 2008). Based on this finding, the present research includes more than one chemotype of *O. tenuiflorum*, followed by different samples of *O. gratissimum* and *O. basilicum*. While preparing polymorphic tables, various AA arrangements were studied in detail to find the similarities and differences in their arrangements in various *Ocimum* species. These results are cross checked by HPLC testing during the chemical analysis process. Further details of chemical test results can be found in chapter 7.

There is limited data available regarding eugenol production in various *Ocimum* species and *OMT* enzyme. The *EOMT* enzyme level is highest in *O. gratissimum* species, where abundant eugenol is produced. No production was, however, noticed for methyl eugenol, which suggests inactivation of the *EOMT* enzyme (Anand *et al.*, 2016).

By silencing the relevant gene or overexpressing the *EOMT* gene, the enzyme's exact roles can be confirmed in future by using this research findings.

8.2 Future work

8.2.1 Protein expression and *OMT* enzyme assay

Previous studies have revealed that *O. basilicum* glandular trichomes express two *OMT* enzymes, one specific to eugenol and another, similar one specific to chavicol (Gang *et al.*, 2002). An activity ratio of these two *OMT* enzymes is not constant across various sweet basil chemotypes (Lewinsohn *et al.*, 2000; Gang *et al.*, 2001). Based on these findings, the present research designed a bacterial transformation study by using *OMT* sequences from different *Ocimum* species including *O. tenuiflorum*, *O. gratissimum*, *O. basilicum* (*EOMT*) and *O. basilicum* (*CVOMT*). By designing a protein expression study, the presence of these two similar enzymes along with some factors affecting *OMT* activities could also be identified. If noticeable differences are observed within these two *OMT* enzymes, further analysis can be carried out to discern if these enzymes are of completely different forms from each other, or whether they overlap on certain substrates. A complete protein expression study of *EOMT* and *CVOMT* enzymes within different *Ocimum* species can certainly answer these questions. Previous studies have already proved that *OMT* enzymes are highly expressed in young leaves which have large amounts of glandular trichomes. The mRNA transcript level is high in these trichomes, and specific enzyme activity is consequently also high. Before cells begin to expand, the majority of glands reach to maturity and *OMT* enzyme activity has started decreasing (Gang *et al.*, 2001). This finding can be used to check the enzyme assay results. The qPCR technique can be useful in analysing *OMT* transcript levels to check *O-methyltransferase* assay results. The final result can be achieved by finding the characterisation of *EOMT* and *CVOMT* proteins as *EOMT* is specific for eugenol while *CVOMT* has similar levels of activity with chavicol, although it can also be active with chavicol and eugenol respectively at comparatively low levels. The separate activity of *EOMT* and *CVOMT* enzymes can be further analysed in different *Ocimum* species. The molecular docking approach can also be used to model the interaction between a small molecule and a protein at the atomic level, which allow to characterize the behavior of small molecules in the binding site of target proteins as well as to elucidate fundamental biochemical processes (McConkey *et al.*, 2002).

8.2.2 *Agrobacterium* transformation

By nature, *Agrobacterium* species act as plant pathogens, but they have been used as a DNA carrier for more than three decades in the field of genetic plant engineering. By using an improved biotechnological technique, further steps in plant transformation can be designed. The major improvements can be seen in the conditions of regeneration and alteration when plant tissue cultures are transformed. However, *Agrobacterium*-mediated transformation is highly complex, including genetic manipulation of the bacterium as well as the host plant cell. Both *Arabidopsis* and *Ocimum* can be used as host plants. The *pK2GW7* vector was used to prepare the final construct of Gateway cloning products. This *Agrobacterium*-harbouring vector contains four *Ocimum* OMT cloning products including *O. basilicum* (EOMT), *O. basilicum* (CVOMT), *O. tenuiflorum* and *O. gratissimum*, which can be used in future to transform *Arabidopsis* plants using the floral dip method (Zhang *et al.*, 2006), and the *Ocimum* plant using both the floral dip and tissue culture methods. The plants can be grown in temperature-controlled growth chambers, and next generation seeds can be collected to check the success of the transformation.

8.2.3 *Arabidopsis* and *Ocimum* plant transformation

Compared to other plant species, *Arabidopsis thaliana* plant transformation is simple and quick. It requires no regeneration or plant tissue culture. *Arabidopsis* flowers can be soaked in an *Agrobacterium tumefaciens* solution along with sucrose and some surfactant solution. The success rate of this method is quite high, as both young and mature flower buds can produce transformed offspring quickly. The repetition of this process can increase its success rate as well as produce new, transformed generations. The transformation rate can also be increased by covering plants to retain humidity levels. Further attempts including position cloning and T-DNA gene tagging can be designed based on this modified method (Gelvin, 2003).

In 2002, Deschamps and Simon (2002) first carried out the *Ocimum* transformation using the *Agrobacterium*-mediated transformation system. GV3101 and EHA105 *Agrobacterium* strains along with β -Glucuronidase (GUS) transient expression were used to obtain positive results. Using similar approach, infected leaf segments can be regenerated using MS media with suitable plant growth hormones. New plants regenerate after in vitro cultivation. The transformation can be confirmed through PCR and Southern blot analysis. Transgene

inheritance can be further checked by colorimetric assay and by PCR of the second generation of transgenic plants. The phenotypic characteristic of T1 plants are identical and accumulate high concentrations of target secondary metabolites.

By using these tools, both *Arabidopsis* and *Ocimum* plants can be transformed using *OMT* genes to check their activity with regard to the production of the relevant phenylpropanoids. For example, according to Anand *et al.* (2016), abundant eugenol production has been recorded in *O. gratissimum*, although no any traces of methyl eugenol have been found. By inserting the *EOMT* enzyme into *O. gratissimum*, the conversion of eugenol in its methyl derivative can be tested. Similarly, some chavicol production has been recorded in *O. gratissimum*. By inserting the *CVOMT* enzyme, estragole production can also be tested.

8.2.4 Examine the expression of key biosynthetic genes under environmental and plant growth conditions which affect ME levels in varieties of Tulsi

As Tulsi is a sacred plant in the Hindu community, it is present in all the locations around the world to where this community has migrated. Extensive research on large tree populations regarding the migration and environmental factors that affect the levels of genetic diversity (Storfer *et al.*, 2007) has been conducted. Similar studies could be carried out by growing identical plants from seeds from various regions followed by checking their *OMT* genetic activity and chemical contents. This study can be further extended by controlling different parameters such as temperature, plant development stages and harvesting methods.

Reference

- Abdelhamid, Y., Lac, N.D., Bs, M.S., 2002. Unani Medicine, Fourth Edi. ed, Textbook of Natural Medicine. Elsevier. doi:10.1016/B978-1-4377-2333-5.00049-3.
- Abdool Karim SS, Ziqubu-Page TT, Arendse R., 1994. Bridging the gap: potential for a health care partnership between African traditional healers and biomedical personnel in South Africa. *S Afr Med J*.;84: s1-s16.
- Abrc, 2012. Methods used by the Arabidopsis Biological Resource Center. Arabidopsis.
- Adams, M.D., 2000. The Genome Sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195. doi:10.1126/science.287.5461.2185.
- Ahmad, S. and Spoor, W., 1999. Effects of NAA and BAP on Callus Culture and Plant Regeneration in Curly Kale (*Brassica oleracea* L.). *Pakistan Journal of Biological Sciences*, 2: 109-112.
- Ali, M., 1990. Ayurvedic drugs in unani materia medica. *Ancient Science Life* 4, 191–201. doi:ASL-9-191 [pii].
- Ali, S.M., Khan, A.A., Ahmed, I., Musaddiq, M., Ahmed, K.S., Polasa, H., Rao, L.V., Habibullah, C.M., Sechi, L.A., Ahmed, N., 2005. Antimicrobial activities of Eugenol and Cinnamaldehyde against the human gastric pathogen *Helicobacter pylori*. *Annals of Clinical Microbiology and Antimicrobials* 4. doi:10.1186/1476-0711-4-20.
- Alter, J.S., 2008. Rethinking the history of medicine in Asia: Hakim Mohammed Said and the society for the promotion of Eastern Medicine. *Journal of Asian Studies*. doi:10.1017/S0021911808001745.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*. doi:10.1093/nar/25.17.3389.
- Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*. doi:10.1016/S1055-7903(03)00208-2.
- Alves, A., Quecini, V. and Vieira, M., 1999. Plant transformation: advances and perspectives *Scientia Agricola*, 56, 1-8.
- Anand, A., Jayaramaiah, R.H., *et al.*, 2016. Comparative functional characterization of eugenol synthase from four different *Ocimum* species: Implications on eugenol accumulation. *Biochimica et Biophysica Acta - Proteins and Proteomics* 1864, 1539–1547. doi:10.1016/j.bbapap.2016.08.004.
- Anthony-Cahill S. J. , Griffith M. C. , Noren C. J. , Suich D. J. , Schultz P. G., 1989. Site-specific mutagenesis with unnatural amino acids. *Trends Biochem. Sci.*;14:400–403.
- Arteca, R., 1996. *Plant Growth Substances: Principles and Applications*. New York: Chapman & Hall.

Ashburner, G.R., Thompson, W.K., Burch, J.M., 1993. Effect of α -naphthaleneacetic acid and sucrose levels on the development of cultured embryos of coconut. *Plant Cell, Tissue and Organ Culture* 35, 157–163. doi:10.1007/BF00032965.

Bai, J., Baldwin, E., Liao, H.L., Zhao, W., Kostenyuk, I., Burns, J., Irely, M., 2013. Extraction of DNA from orange juice, and detection of bacterium *Candidatus liberibacter asiaticus* by real-time PCR. *Journal of Agricultural and Food Chemistry* 61, 9339–9346. doi:10.1021/jf402364y.

Bakkali, F., Averbeck, S., Averbeck, D., Idaomar, M., 2008. Biological effects of essential oils - A review. *Food and Chemical Toxicology*. doi:10.1016/j.fct.2007.09.106.

Bandaranayake, W.M., 2002. Bioactivities, bioactive compounds and chemical constituents of mangrove plants. *Wetlands Ecology and Management*. doi:10.1023/A:1021397624349.

Bandaranayake, W.M., 2006. Quality Control, Screening, Toxicity, and Regulation of Herbal Drugs, in: *Modern Phytomedicine: Turning Medicinal Plants into Drugs*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 25–57. doi:10.1002/9783527609987.ch2.

Banerjee, M., 2002. Power, culture and medicine: Ayurvedic pharmaceuticals in the modern market. *Contributions to Indian Sociology* 36, 435–467. doi:10.1177/006996670203600301.

Baneyx F., 1999. Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology*, 10(5):411–421. 10.1016/S0958-1669(99)00003-8.

Bannasch D, Mehrle A, Glatting KH, Pepperkok R, Poustka A, Wiemann S, 2004. LIFEdb: a database for functional genomics experiments integrating information from external sources, and serving as a sample tracking system. *Nucleic Acids Research*, 32: D505-D508. 10.1093/nar/gkh022.

Banu, L. and Bari, M. 2007. Protocol Establishment for Multiplication and Regeneration of *Ocimum sanctum* Linn. An Important Medicinal Plant with High Religious Value in Bangladesh. *Journal of Plant Sciences*, 2: 530-537.

Baxevanis, A.D., 2004. An overview of gene identification: approaches, strategies, and considerations. *Current protocols in bioinformatics / editorial board*, Chapter 4, Unit4.1. doi:10.1002/0471250953.bi0401s6.

Behere, P., Das, A., Yadav, R., Behere, A., 2013. Ayurvedic concepts related to psychotherapy. *Indian Journal of Psychiatry* 55, 310. doi:10.4103/0019-5545.105556.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A., Wheeler, D.L., 2000. GenBank. *Nucleic Acids Res.* 2000.Jan.1.;28.(1):15-8. 28, 15–18.

Benzie, I.F.F., Watchel-Galor, S., 2011. Herbal Medicine: an introduction to its history, usage, regulation, current trends, and research needs, *Herbal Medicine: biomolecular and clinical aspects*. CRC Press/Taylor & Francis. doi:doi:10.1201/b10787-2.

Bernard, P., Gabarit, P., Bahassi, E.M., Couturier, M., 1994. Positive-selection vectors using the F plasmid *ccdB* killer gene. *Gene* 148, 71–74. doi:10.1016/0378-1119(94)90235-6.

- Bevan, M., 1984. Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acids Research* 12, 8711–8721. doi:10.1093/nar/12.22.8711.
- Bhamra, S., 2016. Investigating the Use and Identity of Traditional Herbal Remedies Amongst South Asian Communities Using Surveys and Biomolecular Techniques. Published Thesis (PhD), Leicester, De Montfort University.
- Bhamra, S., Heinrich, M., Howard, C., Johnson, M. and Slater, A., 2015. DNA authentication of tulsi (*Ocimum tenuiflorum*) using the nuclear ribosomal internal transcribed spacer (ITS) and the chloroplast intergenic spacer trnH-psbA. *Planta Med* 2015; 81 - PW_20.
- Bhatia, J.C., Vir, D., Timmappaya, A., Chuttani, C.S., 1975. Traditional healers and modern medicine. *Social Science & Medicine* (1967) 9, 15–21. doi:10.1016/0037-7856(75)90153-5.
- Bhattacharya, S., 2016. Chapter 3 – Cultivation of Essential Oils, in: *Essential Oils in Food Preservation, Flavor and Safety*. pp. 19–29. doi:10.1016/B978-0-12-416641-7.00003-1.
- Biswas N and Biswas A, 2005. Evaluation of some leaf dusts as grain protectant against rice weevil *Sitophilus oryzae* (Linn.) *Environ Ecol*; 23:485–8.
- Blaikie, C., 2013. Currents of Tradition in Sowa Rigpa Pharmacy. *East Asian Science, Technology and Society* 7, 425–451. doi:10.1215/18752160-2332223.
- Blaikie, C., Craig, S., Gerke, B., Hofer, T., 2015. Coproducing Efficacious Medicines Collaborative Event Ethnography with Himalayan and Tibetan Sowa Rigpa Practitioners. *Current Anthropology* 56, 178–204. doi:10.1086/680464.
- Bloom K., 2005. Chromosome segregation: seeing is believing. *Curr. Biol.* 15:R500–503.
- Booker A, Jalil B, Frommenwiler D, Reich E, Zhai L, Kulic Z, Heinrich M. The authenticity and quality of *Rhodiola rosea* products. *Phytomedicine* 2016; 23: 754-762.
- Borchert P, Miller JA, Miller EC, Shires TK. 1973. 10-Hydroxysafrole, a proximate carcinogenic metabolite of safrole in the rat and mouse. *Cancer Res* 33:590–600.
- Breda, A., Valadares, N.F., Souza, O.N. de, Garratt, R.C., 2007. Protein Structure, Modelling and Applications. *Bioinformatics in Tropical Disease Research: A Practical and Case-Study Approach* 1–41.
- Bremer, B., 2007. Linnaeus' sexual system and flowering plant phylogeny. *Nordic Journal of Botany* 25, 5–6. doi:10.1111/j.0107-055X.2007.00098_12.x.
- British Pharmacopoeia Commission, 2016. St. John's Wort, Ph Eur monograph 1438. British Pharmacopoeia. London, England, The Stationary Office.
- Brooks, J., 2002. Ayurveda, in: *Handbook of Complementary and Alternative Therapies in Mental Health*. pp. 453–471. doi:10.1016/B978-012638281-5/50022-X.
- Burt, S., 2004. Essential oils: their antibacterial properties and potential applications in foods—a review. *International journal of food microbiology* 94, 223–253. doi:10.1016/j.ijfoodmicro.2004.03.022.

Cabeen M.T. , Jacobs-Wagner C., 2005. Bacterial cell shape. *Nat. Rev. Microbiol*;3:601–610.

Calabrese, R.L., 2014. Channeling the Central Dogma. *Neuron*.
doi:10.1016/j.neuron.2014.05.008.

Cappers, R., & Bekker, R., 2013. A manual for the identification of plant seeds and fruits. (Groningen Archaeological Series; Vol. 23). Barkhuis Publishing & Groningen University Library.

Carvajal-Vallejos, P.K., Campos, A., Fuentes-Prior, P., Villalobos, E., Almeida, A.M., Barberà, E., Torné, J.M., Santos, M., 2007. Purification and in vitro refolding of maize chloroplast transglutaminase over-expressed in *Escherichia coli*. *Biotechnology Letters* 29, 1255–1262. doi:10.1007/s10529-007-9377-7.

Cavalier-Smith, T., 1981. Eukaryote kingdoms: Seven or nine? *BioSystems* 14, 461–481.
doi:10.1016/0303-2647(81)90050-2

Chang, S.S., Park, S.K., Kim, B.C., Kang, B.J., Kim, D.U., Nam, H.G., 1994. Stable genetic transformation of *Arabidopsis thaliana* by *Agrobacterium* inoculation in planta. *The Plant Journal* 5, 551–558. doi:10.1046/j.1365-313X.1994.5040551.x.

Chase, M.W. Salamin, N. Wilkinson, M. Dunwell, J.M. Kesanakurthi, R. P. Haidar, N. Savolainen, V., 2005. Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society B*, 360, p1889-1895.

Chase, M.W., Hills, H.H., 1991. Silica Gel: An Ideal Material for Field Preservation of Leaf Samples for DNA Studies. *Taxon* 40, 215. doi:10.2307/1222975.

Chaudhary, D. & Dahal, R., 2017. DNA Bar-Code for Identification of Microbial Communities: A Mini-Review. *EC Microbiology Research Article*. 7. 219-224.

Chen, R., Dong, J., Cui, X., Wang, W., Yasmeen, A., Deng, Y., Zeng, X., Tang, Z., 2012. DNA based identification of medicinal materials in Chinese patent medicines. *Scientific Reports* 2. doi:10.1038/srep00958.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y., Leon, C., 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* 5.
doi:10.1371/journal.pone.0008613.

Cheng, X., Chen, X., Su, X., Zhao, H., Han, M., Bo, C., Xu, J., Bai, H., Ning, K., 2014. DNA Extraction Protocol for Biological Ingredient Analysis of Liuwei Dihuang Wan. *Genomics, Proteomics and Bioinformatics* 12, 137–143. doi:10.1016/j.gpb.2014.03.002.

Chevallier, A., 2001. Encyclopedia of Medicinal Plants. *Encyclopedia of Medicinal Plants* 18–27, 211–212.

Chopra, A., Doiphode, V.V., 2002. Ayurvedic medicine: Core concept, therapeutic principles, and current relevance. *Medical Clinics of North America*. doi:10.1016/S0025-7125(03)00073-7.

- Choudhuri, S., 2014. Chapter 9 - Phylogenetic Analysis*, in: *Bioinformatics for Beginners*. Academic Press, pp. 209–218. doi:<https://doi.org/10.1016/B978-0-12-410471-6.00009-8>.
- Chowdhury, T., Mandal, A., Roy, S.C., De Sarker, D., 2017. Diversity of the genus *Ocimum* (Lamiaceae) through morpho-molecular (RAPD) and chemical (GC–MS) analysis. *Journal of Genetic Engineering and Biotechnology* 15, 275–286. doi:10.1016/j.jgeb.2016.12.004.
- Christenhusz MJM, Byng JW. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa* 261:201–217.
- Christina, V. & Annamalai, A., 2014. Nucleotide based validation of *Ocimum* species by evaluating three candidate barcodes of the chloroplast region in *Molecular ecology resources*, vol. 14, no. 1, pp. 60-68.
- Chudakova, T., 2017. Plant matters: Buddhist medicine and economies of attention in postsocialist Siberia. *American Ethnologist* 44, 341–354. doi:10.1111/amet.12483.
- Clark, S.E., Williams, R.W., Meyerowitz, E.M., 1997. The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* 89, 575–585. doi:10.1016/S0092-8674(00)80239-1.
- Clay, N.K., Nelson, T., 2002. VH1, a provascular cell-specific receptor kinase that influences leaf cell patterns in *Arabidopsis*. *The Plant cell* 14, 2707–22. doi:10.1105/tpc.005884.
- Cline, J., 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research* 24, 3546–3551. doi:10.1093/nar/24.18.3546.
- Cohen, M. (2014). Tulsi - *Ocimum sanctum*: A herb for all reasons. *Journal of Ayurveda and Integrative Medicine*, 5(4), 251. <https://doi.org/10.4103/0975-9476.146554>.
- Committee, on H.M.P., 2011. Assessment report on *Syzygium aromaticum* (L.) Merrill et L. M. Perry, flos and *Syzygium aromaticum* (L.) Merrill et. European Medicines Agency (EMA) 44, 26.
- Conboy, L., Kaptchuk, T.J., Eisenberg, D.M., Gottlieb, B., Acevedo-Garcia, D., 2007. The relationship between social factors and attitudes toward conventional and CAM practitioners. *Complementary Therapies in Clinical Practice* 13, 146–157. doi:10.1016/j.ctcp.2006.12.003.
- Conn, B.J., 2014. *Ocimum* L. (Lamiaceae) in Australia and Papua New Guinea. *Telopea*. doi:10.7751/telopea20147790.
- Cortés-Rojas, D.F., de Souza, C.R.F., Oliveira, W.P., 2014. Clove (*Syzygium aromaticum*): A precious spice. *Asian Pacific Journal of Tropical Biomedicine*. doi:10.1016/S2221-1691(14)60215-X.
- Cotton C. M., 1996. *Ethnobotany: Principles and Applications*. Chichester, England: Wiley.
- Crick, F.H.C., 1970. Central Dogma of Molecular Biology. *Nature*. doi:10.1038/227561a0.
- Croteau, R., Kutchan, T.M., Lewis, N.G., 2000. Natural products (secondary metabolites), in: *Biochemistry Molecular Biology of Plants*. pp. 1250–1318. doi:10.1016/B978-0-08-088504-9.00018-0.

CRISM (Centre for Research on Indian Systems of Medicine), 2008.
<http://www.crism.net/overview.htm>.

Cubero, O.F., Crespo, A., Fatehi, J., Bridge, P.D., 1999. DNA extraction and PCR amplification method suitable for fresh, herbarium-stored, lichenized, and other fungi. *Plant Systematics and Evolution* 216, 243–249. doi:10.1007/BF01084401.

Datlof, E.M., Amend, A.S., Earl, K., Hayward, J., Morden, C.W., Wade, R., Zahn, G., Hynson, N.A., 2017. Uncovering unseen fungal diversity from plant DNA banks. *PeerJ* 5, e3730. doi:10.7717/peerj.3730.

de Boer, H.J., Ichim, M.C., Newmaster, S.G., 2015. DNA Barcoding and Pharmacovigilance of Herbal Medicines. *Drug Safety* 38, 611–620. doi:10.1007/s40264-015-0306-8.

Dejonghe, W., Russinova, E., 2014. Target identification strategies in plant chemical biology. *Front Plant Sci* 5, 352. doi:10.3389/fpls.2014.00352.

Delves P. J. , Roitt I. M., 2000. The immune system. Second of two parts. *N. Engl. J. Med.* 343:108–117.

Deoxyribonucleic acid (DNA) based identification techniques for herbal drugs. *British Pharmacopoeia Appendix XI V*, 2017.

Deschamps, C., Simon, J.E., 2002. *Agrobacterium tumefaciens*-mediated transformation of *Ocimum basilicum* and *O. citriodorum*. *Plant Cell Reports* 21, 359–364. doi:10.1007/s00299-002-0526-0.

Dexter, R., Qualley, A., Kish, C.M., Ma, C.J., Koeduka, T., Nagegowda, D.A., Dudareva, N., Pichersky, E., Clark, D., 2007. Characterization of a petunia acetyltransferase involved in the biosynthesis of the floral volatile isoeugenol. *Plant Journal* 49, 265–275. doi:10.1111/j.1365-3113X.2006.02954.x.

Dhanapal, A.P., Govindaraj, M., 2015. Unlimited thirst for genome sequencing, data interpretation, and database usage in genomic era: The road towards fast-track crop plant improvement. *Genetics Research International*. doi:10.1155/2015/684321.

Dholakia, B.B., Puneekar, S.A., Gade, W.N., Thulasiram, H.V., Giri, A.P., 2016. Comparative functional characterization of eugenol synthase from four different *Ocimum* species: Implications on eugenol accumulation. *Biochimica et Biophysica Acta - Proteins and Proteomics* 1864, 1539–1547. doi:10.1016/j.bbapap.2016.08.004.

Díaz-Maroto, M., Sánchez Palomo, E., Castro, L., González Viñas, M. and Pérez-Coello, M. (2004). Changes produced in the aroma compounds and structural integrity of basil (*Ocimum basilicum*L.) during drying. *Journal of the Science of Food and Agriculture*, 84(15), pp.2070-2076.

DNA barcoding as a tool for botanical identification. *British Pharmacopoeia Supplementary Chapter SC VII D*, 2017.

Dode, L.B., Bobrowski, V.L., Jacira, E., Braga, B., Seixas, K., Schuch, M.W., 2003. In vitro propagation of *Ocimum basilicum* L. (Lamiaceae). *Acta Scientiarum. Biological Sciences* 25, 435–437. doi:10.4025/actascibiolsci.v25i2.2034.

Dörnenburg, H., Knorr, D., 1995. Strategies for the improvement of secondary metabolite production in plant cell cultures. *Enzyme and Microbial Technology*. doi:10.1016/0141-0229(94)00108-4.

Drábková, L., Kirschner, J., Vlček, Č., 2002. Comparison of seven DNA extraction and amplification protocols in historical herbarium specimens of juncaceae. *Plant Molecular Biology Reporter* 20, 161–175. doi:10.1007/BF02799431.

Edwards D., Batley J., 2004. Plant bioinformatics: from genome to phenome. *Trends Biotechnol*;22(5):232–237.

EMA, 2014. Public statement on the use of herbal medicinal products containing estragole. Committee on Herbal Medicinal Products (HMPC) 44, 1–9.

Endress, P.K., 2001. Origins of flower morphology. *Journal of Experimental Zoology*. doi:10.1002/jez.1063.

Engelbreton, J., 2002. Culture and complementary therapies. *Complementary Therapies in Nursing and Midwifery* 8, 177–184. doi:10.1054/ctnm.2002.0638.

Ernst E, Pittler MH, Wider B, Boddy K: Oxford Handbook of Complementary Medicine. Oxford, Oxford University Press, 2008.

Erickson, D.L., Reed, E., Ramachandran, P., Bourg, N.A., McShea, W.J., Ottesen, A., 2017. Reconstructing a herbivore's diet using a novel rbcL DNA mini-barcode for plants. *AoB PLANTS*. doi:10.1093/aobpla/plx015.

Escudero, J., Neuhaus, G., Schläppi, M., Hohn, B., 1996. T-DNA transfer in meristematic cells of maize provided with intracellular *Agrobacterium*. *Plant Journal* 10, 355–360. doi:10.1046/j.1365-3113X.1996.10020355.x.

Esfahani, R.E., Moradi, P., 2017. Experimental Paper. The effect of different development stages on the quantity and quality of the essential oil of *Citrus aurantifolia* (Christm.) Swingle in Iran. *Herba Polonica* 63. doi:10.1515/hepo-2017-0003.

Fazekas, A.J., Steeves, R., Newmaster, S.G., 2010. Improving sequencing quality from PCR products containing long mononucleotide repeats. *BioTechniques* 48, 277–283. doi:10.2144/000113369.

Ferrer, J.L., Austin, M.B., Stewart, C., Noel, J.P., 2008. Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiology and Biochemistry*. doi:10.1016/j.plaphy.2007.12.009.

Feuillet, C., Leach, J.E., Rogers, J., Schnable, P.S., Eversole, K., 2011. Crop genome sequencing: Lessons and rationales. *Trends in Plant Science*. doi:10.1016/j.tplants.2010.10.005.

- Fisher, P., Ward, A., 1994. Medicine in Europe: Complementary medicine in Europe. *BMJ* 309, 107. doi:10.1136/bmj.309.6947.107.
- Funk, V., 2003. The Importance of Herbaria. *Plant Science Bulletin* 49, 94–95.
- Gajendrarao, P., Krishnamoorthy, N., Sakthiah, S., Lazar, P., Lee, K.W., 2010. Molecular modeling study on orphan human protein CYP4A22 for identification of potential ligand binding site. *Journal of Molecular Graphics and Modelling* 28, 524–532. doi:10.1016/j.jmgm.2009.11.010.
- Galperin M. Y., 2006. The Molecular Biology Database Collection: update. *Nucl. Acids Res.* 2006;34:D3–D5.
- Gang, D.R., Lavid, N., Zubieta, C., Chen, F., Beuerle, T., Lewinsohn, E., Noel, J.P., Pichersky, E., 2002. Characterization of Phenylpropene O -Methyltransferases from Sweet Basil : Facile Change of Substrate Specificity and Convergent Evolution within a Plant O -Methyltransferase Family. *The Plant Cell* 14, 505–519. doi:10.1105/tpc.010327.
- Gang, D.R., Wang, J., Dudareva, N., Nam, K.H., Simon, J.E., Lewinsohn, E. & Pichersky, E., 2001. An investigation of the storage and biosynthesis of phenylpropenes in sweet basil. *Plant Physiology*, vol. 125, no. 2, pp. 539-555.
- Garfinkel, D.J., Simpson, R.B., Ream, L.W., White, F.F., Gordon, M.P., Nester, E.W., 1981. Genetic analysis of crown gall: Fine structure map of the T-DNA by site-directed mutagenesis. *Cell* 27, 143–153. doi:10.1016/0092-8674(81)90368-8.
- Gartner, B.L., 1995. Plant stems: physiology and functional morphology 125. doi:10.1016/B978-012276460-8/50008-4.
- Gavali, J.B., Pradhan, N., Waghmare, N., 2016. Current Trends in Analytical Methods of Medicinal Plant Drugs. *International Ayurveda Publications* 1, 101–107.
- Gelvin, S.B., 2003. Agrobacterium-Mediated Plant Transformation: the Biology behind the “Gene-Jockeying” Tool. *Microbiology and Molecular Biology Reviews* 67, 16–37. doi:10.1128/MMBR.67.1.16-37.2003.
- Gemeinholzer, B., Dröge, G., Zetzsche, H., Haszprunar, G., Klenk, H.-P., Güntsch, A., Berendsohn, W.G., Wägele, J.-W., 2011. The DNA Bank Network: The Start from a German Initiative. *Biopreservation and Biobanking* 9, 51–55. doi:10.1089/bio.2010.0029.
- Ghosh, P.K., Bhattacharjee, P., 2016. Mathematical modeling of supercritical carbon dioxide extraction of methyl eugenol from tuberose flowers. *Korean Journal of Chemical Engineering* 33, 1681–1691. doi:10.1007/s11814-015-0247-z.
- Gillings, M.R., Westoby, M., 2014. DNA technology and evolution of the Central Dogma. *Trends in Ecology & Evolution* 29, 1–2. doi:10.1016/j.tree.2013.10.001.
- Godiard, L., Sauviac, L., Torii, K.U., Grenon, O., Mangin, B., Grimsley, N.H., Marco, Y., 2003. ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. *Plant Journal* 36, 353–365. doi:10.1046/j.1365-313X.2003.01877.x.

- Gomord, V., Faye, L., 2004. Posttranslational modification of therapeutic proteins in plants. *Current Opinion in Plant Biology*. doi:10.1016/j.pbi.2004.01.015.
- Gonzalez, M.A., Baraloto, C., Engel, J., Mori, S.A., Pétronelli, P., Riéra, B., Roger, A., Thébaud, C., Chave, J., 2009. Identification of amazonian trees with DNA barcodes. *PLoS ONE* 4. doi:10.1371/journal.pone.0007483.
- Government of India. (2007-2012). Report of the Working Group on Agricultural Research and Education for the Eleventh Five Year Plan. Publisher. Length, 458 pages.
- Govindaraj, M., Vetriventhan, M., Srinivasan, M., 2015. Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. *Genetics Research International*. doi:10.1155/2015/431487.
- Grayer, R.J., Kite, G.C., Goldstone, F.J., Bryan, S.E., Paton, A., Putievsky, E., 1996. Intraspecific taxonomy and essential oil chemotypes in sweet basil, *Ocimum basilicum*. *Phytochemistry* 43, 1033–1039. doi:10.1016/S0031-9422(96)00429-3.
- Green M. R., 2005. Eukaryotic transcription activation: right on target. *Mol. Cell*;18:399–402.
- Guidance Manual for Compliance of Indian Pharmacopoeia. 2012. The Indian Pharmacopoeia Commission, Ghaziabad.
- Guidance Manual for Monographs Development of Herbs and Herbal Products including Phytopharmaceutical Drugs. 2016. The Indian Pharmacopoeia Commission, Ghaziaba.
- Gunturu, K.S., Nagarajan, P., McPhedran, P., Goodman, T.R., Hodsdon, M.E., Strout, M.P., 2011. Ayurvedic herbal medicine and lead poisoning. *Journal of Hematology and Oncology* 4. doi:10.1186/1756-8722-4-51.
- Günzburg, W.H., Salmons, B., 1995. Virus vector design in gene therapy. *Molecular Medicine Today*. doi:10.1016/S1357-4310(95)90771-8.
- Gupta, P.D., 2015. Pharmacogenetics, Pharmacogenomics and Ayurgenomics for Personalized Medicine: A Paradigm Shift. *Indian Journal of Pharmaceutical Sciences* 77, 135–141.
- Gupta, V.K., Singh, J., Kumar, R., Bhanot, A., 2011. Pharmacognostic and preliminary phytochemical study of *Ocimum gratissimum* Linn. (Family: Lamiaceae). *Asian Journal of Plant Sciences* 10, 365–369. doi:10.3923/ajps.2011.365.369.
- Gurib-Fakim A., 2006. Medicinal plants: traditions of yesterday and drugs of tomorrow. *Mol Aspects Med.*;27:1-93.
- Hahlbrock K, Scheel D, 1989. Physiology and molecular biology of phenylpropanoid metabolism. *Annu Rev Plant Physiol Plant Mol Biol* 40: 347–369.
- Hahn-Deinstrop, E., 2006. Applied Thin-Layer Chromatography, Applied Thin-Layer Chromatography. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. doi:10.1002/9783527610259.

- Hankey, a, 2001. Ayurvedic physiology and etiology: Ayurvedo Amritanaam. The doshas and their functioning in terms of contemporary biology and physical chemistry. *Journal of alternative and complementary medicine* (New York, N.Y.) 7, 567–74. doi:10.1089/10755530152639792.
- Hansen, G., Chilton, M.D., 1996. “Agrolistic” transformation of plant cells: integration of T-strands generated in planta. *Proc Natl Acad Sci U S A* 93, 14978–14983. doi:10.1073/pnas.93.25.14978.
- Haque I, Bandopadhyay R, Mukhopadhyay K, 2008. An optimised protocol for fast genomic DNA isolation from high secondary metabolite and gum containing plants. *Asian J Plant Sci* 7:304–308. doi:10.3923/ajps.2008.304.308.
- Harris, D.J., 2003. Can you bank on GenBank? *Trends in Ecology and Evolution*. doi:10.1016/S0169-5347(03)00150-2.
- Hartley, J.L., 2006. Cloning technologies for protein expression and purification. *Current Opinion in Biotechnology*. doi:10.1016/j.copbio.2006.06.011.
- Hartmann, T. (2007) From waste products to ecochemicals: Fifty years research of plant secondary metabolism. *Phytochemistry*, 68, 2831-2846. doi:10.1016/j.phytochem.2007.09.017.
- Hayat, N., & Naeem, A., 2014. Corporate Waqf: A Case of Hamdard (Waqf) Pakistan. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2487713>.
- Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* 270, 313–321. doi:10.1098/rspb.2002.2218.
- Hereman S., 1868. *Paxton’s Botanical Dictionary*, Bradbury Evans and Co., London.
- Heywood, V.H., Heenan, P.B., Brummitt, R.K., Culham, A., Seberg, O., 2008. *Flowering plant families of the world*, New Zealand Journal of Botany. Firefly Books, New York. doi:10.1017/S0960428607064335.
- Hirt H, M’Pia B., 2008. *Natural Medicine in the Tropics I: Foundation Text*. Winnenden: Anamed.
- Hodacs, H., 2010. In the field: Exploring nature with Carolus Linnaeus. *Endeavour*. doi:10.1016/j.endeavour.2009.11.001.
- Hodkinson, T.R., Waldren, S., Parnell, J.A.N., Kelleher, C.T., Salamin, K., Salamin, N., 2007. DNA banking for plant breeding, biotechnology and biodiversity evaluation. *Journal of Plant Research* 120, 17–29. doi:10.1007/s10265-006-0059-7.
- Hollingsworth, M.L., Andra Clark, A., Forrest, L.L., Richardson, J., Pennington, R.T., Long, D.G., Cowan, R., Chase, M.W., Gaudeul, M., Hollingsworth, P.M., 2009. Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources* 9, 439–457. doi:10.1111/j.1755-0998.2008.02439.x.

- Hollingsworth, P.M., Graham, S.W., Little, D.P., 2011. Choosing and using a plant DNA barcode. *PLoS ONE* 6. doi:10.1371/journal.pone.0019254.
- Hoppe, J., 2011. Curcuma: Turmeric. *Medical Herbalism* 4, 3–5.
- Hosbach, I., Neeb, G., Hager, S., Kirchhoff, S., Kirsch, B., 2003. Traditional Chinese herbal medicine. *Pharmacy world science PWS* 58, 103–112.
- Howard C, Bremner PD, Fowler MR and Slater A., 2008. Molecular identification of St John's Wort by PCR amplification of the ITS1 region: implications for medicinal plant identification. *Journal of Pharmacy and Pharmacology; Supp 1*: A10 – 24.
- Howard C., Smith, S., Bremner, P., Fowler, M., Scott, N., and Slater, A., 2011. The design of DNA barcode-specific PCR primers for medicinal plant authentication. *Planta Medica* 77(12). DOI:10.1055/s-0031-1282165
- Howard, C., 2010. The development of Deoxyribonucleic Acid (DNA) based methods for the identification and authentication of. Published Thesis (PhD), Leicester, De Montfort University.
- Huchelmann, A., Boutry, M., Hachez, C., 2017. Plant glandular trichomes: natural cell factories of high biotechnological interest. *Plant Physiology* pp.00727.2017. doi:10.1104/pp.17.00727.
- Ibrahim, R.K., Bruneau, A., and Bantignies, B., 1998. Plant O-methyltransferases. Molecular analysis, common signature and classification. *Plant Mol. Biol.* 36, 1–10.
- Ii, T., 1989. Restriction enzyme digestion. *Electrophoresis* 2–4. doi:10.1371/journal.pone.0124783.
- Ikeda K, Nakazawa H, Shimo-Oka A, Ishio K, Miyata S, Hosokawa Y, Matsumura S, Masuhara H, Belloncik S, Alain R, 2006. Immobilization of diverse foreign proteins in viral polyhedra and potential application for protein microarrays. *Proteomics*, 6(1):54–66. 10.1002/pmic.200500022.
- International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi:10.1038/35057062.
- International Organization for Migration, 2013. World Migration Report 2013. Migrant Well-Being and Development, World Migration Report. doi:ISBN 9789290686682/ISSN 15615502.
- Jackson, R.J., Hellen, C.U. and Pestova, T.V., 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell. Biol.*, 11, 113–127.
- Jacob, W.F., Santer, M. and Dahlberg, A.E., 1987. A single base change in the Shine-Dalgarno region of 16S rRNA of *Escherichia coli* affects translation of many proteins. *Proc. Natl Acad. Sci. USA*, 84, 4757–4761.

- Jaiswal, Y.S., Williams, L.L., 2017. A glimpse of Ayurveda – The forgotten history and principles of Indian traditional medicine. *Journal of Traditional and Complementary Medicine*. doi:10.1016/j.jtcme.2016.02.002.
- Jamshidi, N., Cohen, M.M., 2017. The Clinical Efficacy and Safety of Tulsi in Humans: A Systematic Review of the Literature. *Evidence-based Complementary and Alternative Medicine*. doi:10.1155/2017/9217567.
- Janin J. , Chothia C., 1980. Packing of alpha-helices onto beta-pleated sheets and the anatomy of alpha/beta proteins. *J. Mol. Biol.*;143:95–128.
- Jiang, Y., Ninwichian, P., Liu, S., Zhang, J., Kucuktas, H., Sun, F., Kaltenboeck, L., Sun, L., Bao, L., Liu, Z., 2013. Generation of Physical Map Contig-Specific Sequences Useful for Whole Genome Sequence Scaffolding. *PLoS ONE* 8. doi:10.1371/journal.pone.0078872.
- Jiaxiang S., 1997. Introduction to the Chinese Materia Medica, China Pharmaceutical University, 24 Tong Jia Xiang, Nanjing 210009. The People's Republic of china, In UNDP.
- Jirovetz, L., Buchbauer, G., Shafi, M.P., Kaniampady, M., 2003. Chemotaxonomical analysis of the essential oil aroma compounds of four different *Ocimum* species from southern India. *European Food Research and Technology* 217, 120–124. doi:10.1007/s00217-003-0708-1.
- Jobes, D.V., Hurley, D.L., Thien, L.B., 1995. Plant DNA isolation: A method to efficiently remove polyphenolics, polysaccharides, and RNA. *Taxon* 44, 379–386. doi:10.2307/1223408.
- Joly, A., Müller, H., Goëau, H., Glotin, H., Spampinato, C., Rauber, A., Bonnet, P., Vellinga, W.P., Fisher, B., 2014. LifeCLEF: Multimedia life species identification, in: CEUR Workshop Proceedings. CEUR-WS, pp. 7–13. doi:10.1007/978-3-319-11382-1_20.
- Joseph, L.J., Bhartiya, U.S., Raut, Y.S., Hawaldar, R.W., Nayak, Y., Pawar, Y.P., Jambhekar, N.A., Rajan, M.G.R., 2011. Radioprotective Effect of *Ocimum sanctum* and Amifostine on the Salivary Gland of Rats After Therapeutic Radioiodine Exposure. *Cancer Biotherapy & Radiopharmaceuticals* 26, 737–743. doi:10.1089/cbr.2011.1014.
- Joshi, K., Ghodke, Y., Patwardhan, B., 2011. Traditional medicine to modern pharmacogenomics: Ayurveda Prakriti type and CYP2C19 gene polymorphism associated with the metabolic variability. *Evidence-based Complementary and Alternative Medicine* 2011. doi:10.1093/ecam/nep206.
- Kawabe, A., Nukii, H., Furihata, H.Y., 2018. Exploring the history of chloroplast capture in arabis using whole chloroplast genome sequencing. *International Journal of Molecular Sciences* 19. doi:10.3390/ijms19020602.
- Kazi T, Hussain N, Bremner P, Slater A and Howard C., 2013. The application of a DNA based identification technique to over the counter herbal medicines. *Fitoterapia* 87: 27–30.
- Kececiloglu, J.D., Lenhof, H.-P., Mehlhorn, K., Mutzel, P., Reinert, K., Vingron, M., 2000. A polyhedral approach to sequence alignment problems. *Discrete Applied Mathematics* 104, 143–186. doi:10.1016/S0166-218X(00)00194-3.

- Kelm M, Nair M, Strasburg G, DeWitt D., 2000. Antioxidant and cyclooxygenase inhibitory phenolic compounds from *Ocimum sanctum* Linn. *Phytomedicine*; 7:7–13.
- Kerk, N. M., & Sussex, I. M., 2012. Roots and Root Systems. In eLS. Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470015902.a0002058.pub2>.
- Khaki, A.A., Khaki, A., Ahmadi-Ashtiani, H.R., Rastegar, H., Rezazadeh, S., Babazadeh, D., Zahedi, A., Ghanbari, Z., 2010. Treatment effects of ginger rhizome & extract of carrot seed on diabetic nephropathy in rat. *Journal of Medicinal Plants* 9, 75–80.
- Khalil, A.A., Rahman, U.U., Khan, M.R., Sahar, A., Mehmood, T., Khan, M., 2017. Essential oil eugenol: Sources, extraction techniques and nutraceutical perspectives. *RSC Advances*. doi:10.1039/c7ra04803c.
- Khan, S., Fahim, N., Singh, P., Rahman, L.U., 2015. *Agrobacterium tumefaciens* mediated genetic transformation of *Ocimum gratissimum*: A medicinally important crop. *Industrial Crops and Products* 71, 138–146. doi:10.1016/j.indcrop.2015.03.080.
- Khan, S., Ullah, M.W., Siddique, R., Nabi, G., Manan, S., Yousaf, M., Hou, H., 2016. Role of recombinant DNA technology to improve life. *International Journal of Genomics*. doi:10.1155/2016/2405954.
- Kibler, 2014. Establishment of an efficient protocol for the micropropagation of Holy Basil (*Ocimum sanctum* L.). Thesis. The University of Guelph, Ontario, Canada.
- Kim, Y., Lee, G., Jeon, E., Sohn, E.J., Lee, Y., Kang, H., Lee, D.W., Kim, D.H., Hwang, I., 2014. The immediate upstream region of the 5'-UTR from the AUG start codon has a pronounced effect on the translational efficiency in *Arabidopsis thaliana*. *Nucleic Acids Research* 42, 485–498. doi:10.1093/nar/gkt864.
- Klischies, M., Stockigt, J., and Zenk, M.H., 1975. Biosynthesis of the allylphenols eugenol and methyleugenol in *Ocimum basilicum* L. *Chem. Commun.* 21, 879–880.
- Kochhar, A., Sharma, N., Sachdeva, R., 2009. Effect of supplementation of Tulsi (*Ocimum sanctum*) and Neem (*Azadirachta indica*) leaf powder on diabetic symptoms, anthropometric parameters and blood pressure of non insulin dependent male diabetics. *Studies on Ethno-Medicine* 3, 5–9.
- Kong JM, Goh NK, Chia LS, Chia TF, 2003. Recent advances in traditional plant drugs and orchids. *Acta Pharmacol Sin.*;24:7-21.
- Kousik, D.M., Baldev, K., 2012. A review on therapeutic uses of *Ocimum sanctum* Linn (TULSI) with its pharmacological actions. *International Journal of Research in Ayurveda and Pharmacy*. doi:10.7897/2277-4343.03512.
- Kress, W.J., Erickson, D.L., 2008. DNA barcodes: Genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences* 105, 2761–2762. doi:10.1073/pnas.0800476105.
- Kubinyi, H., 1997. QSAR and 3D QSAR in drug design. *Theory Methods and Applications - Chapter: Molecular Superposition for Rational Design*, page 200–201.

- Kurup, P.N.V., 2003. Ayurveda is not merely a traditional but scientific medicine. International Ayurveda Foundation.
- Labra, M., Miele, M., Ledda, B., Grassi, F., Mazzei, M., Sala, F., 2004. Morphological characterization, essential oil composition and DNA genotyping of *Ocimum basilicum* L. cultivars. *Plant Science* 167, 725–731. doi:10.1016/j.plantsci.2004.04.026.
- Lachowicz, K.J., Jones, G.P., Briggs, D.R., Bienvenu, F.E., Palmer, M.V., Mishra, V., Hunter, M.M., 1997. Characteristics of Plants and Plant Extracts from Five Varieties of Basil (*Ocimum basilicum* L.) Grown in Australia. *Journal of Agricultural and Food Chemistry* 45, 2660–2665. doi:10.1021/jf960791h.
- Laekeman, G.M., van Hoof, L., Haemers, A., Berghe, D.A.V., Herman, A.G., Vlietinck, A.J., 1990. Eugenol a valuable compound for in vitro experimental research and worthwhile for further in vivo investigation. *Phytotherapy Research* 4, 90–96. doi:10.1002/ptr.2650040304.
- Lafitte, G., 2015. Maternal mortality in Tibet: Sciences of healing, sowa rigpa and community health workers. Maternal mortality: Risk factors, anthropological perspectives, prevalence in developing countries and preventive strategies for pregnancy-related deaths.
- Lagrouh, F., Dakka, N., & Bakri, Y. (2017, September 1). The antifungal activity of Moroccan plants and the mechanism of action of secondary metabolites from plants. *Journal de Mycologie Medicale*. Elsevier Masson SAS. <https://doi.org/10.1016/j.mycmed.2017.04.008>.
- Lange, B. M., Wildung, M. R., Stauber, E. J., Sanchez, C., Pouchnik, D., & Croteau, R., 2000. Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. *Proceedings of the National Academy of Sciences* 97, 2934–2939. doi:10.1073/pnas.97.6.2934.
- Lao, L., Xu, L., Xu, S., 2012. Traditional chinese medicine. *ntegrative pediatric oncology* 125–135.
- Latchman, D.S., 2008. CHAPTER 1 – DNA SEQUENCES, TRANSCRIPTION FACTORS AND CHROMATIN STRUCTURE, in: *Eukaryotic Transcription Factors*. pp. 1–28. doi:10.1016/B978-012373983-4.50009-2.
- Lawrence C.J., Dong Q., Polacco M.L., Seigfried T.E., Brendel V., 2004. MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res.*;32(Suppl. 1):D393–D397.
- Lawrence, B.M., 1988. A further examination of the variation of *Ocimum basilicum* L. *Developments in Food Science* 18, 161–170.
- Lee, H.S., Im, W., 2012. Identification of ligand templates using local structure alignment for structure-based drug design. *Journal of Chemical Information and Modeling* 52, 2784–2795. doi:10.1021/ci300178e.
- Lesk, A. M., 2001. *Introduction to Protein Architecture*. OUP, Oxford.

- Lewinsohn, E., Ziv-Raz, I.I., Dudai, N., Tadmor, Y., Lastochkin, E., Larkov, O., Chaimovitsh, D., Ravid, U., Putievsky, E., Pichersky, E. & Shoham, Y., 2000. Biosynthesis of estragole and methyl-eugenol in sweet basil (*Ocimum basilicum* L.) Developmental and chemotypic association of allylphenol O-methyltransferase activities. *Plant science: an international journal of experimental plant biology*, vol. 160, no. 1, pp. 27-35.
- Li, H., Li, J., Cong, X.H., Duan, Y.B., Li, L., Wei, P.C., Lu, X.Z., Yang, J.B., 2013. A high-throughput, high-quality plant genomic DNA extraction protocol. *Genetics and Molecular Research* 12, 4526–4539. doi:10.4238/2013.October.15.1.
- Li, J., Li, Z., Li, C., Gou, J., Zhang, Y., 2014. Molecular cloning and characterization of an isoflavone 7-O-glucosyltransferase from *Pueraria lobata*. *Plant Cell Reports* 33, 1173–1185. doi:10.1007/s00299-014-1606-7
- Liljas, A., 2013. Translation, in: *Brenner's Encyclopedia of Genetics: Second Edition*. Elsevier Inc., pp. 127–129. doi:10.1016/B978-0-12-374984-0.01564-3.
- Long, E.L., Jenner, P.M., 1963. Esophageal tumours produced in rats by the feeding of dihydrosafrole. *Federation Proceedings Abstracts* 22, 275.
- Luscombe N.M. , Greenbaum D. , Gerstein M., 2001. What is Bioinformatics? A proposed definition and overview of the field. *Method Inform. Med.*;40:346–358.
- Madke, S.S., Cherian, K.J., Badere, R.S., 2014. A modified Murashige and Skoog media for efficient multipleshoot induction in *G. arborea* Roxb. *Journal of Forestry Research* 25, 557–564. doi:10.1007/s11676-014-0449-y.
- Mahajan N, Rawal S, Verma M, Poddar M, Alok S., 2013. A phytopharmacological overview on *Ocimum* species with special emphasis on *Ocimum sanctum*. *Biomed Prev Nutr*; 3: 185–92.
- Maier, T.S., Kuhn, J., Müller, C., 2010. Proposal for field sampling of plants and processing in the lab for environmental metabolic fingerprinting. *Plant Methods* 6. doi:10.1186/1746-4811-6-6.
- Manikandan, P., Vidjaya Letchoumy, P., Prathiba, D., Nagini, S., 2008. Combinatorial chemopreventive effect of *Azadirachta indica* and *Ocimum sanctum* on oxidant-antioxidant status, cell proliferation, apoptosis and angiogenesis in a rat forestomach carcinogenesis model. *Singapore Medical Journal* 49, 814–822.
- Manitto, P., Monti, D., and Gramatica, P., 1974. Biosynthesis of phenylpropanoid compounds. I. Biosynthesis of eugenol in *Ocimum basilicum* L. *J. Chem. Soc. Perkin Trans. I* 14, 1727–1731.
- Mannan, A.A.S.R., Kahvic, M., 2010. Ibn Sina : a tribute. *The gulf journal of oncology* 60–63.
- Mannhold, R., Kubinyi, H., Timmerman, H., 2008. *Structure-based Ligand Design, Structure-based Ligand Design*. wiley. doi:10.1002/9783527612178.

- Marin, K., Krämer, R., Wendisch, V.F., 2007. Amino Acid Biosynthesis - Pathways, Regulation and Metabolic Engineering, Amino Acid Biosynthesis~ Pathways. doi:10.1007/7171_2006_069.
- Martins, P.F., Carmona, C., Martinez, E.L., Sbaite, P., Filho, R.M.I., MacIel, M.R.W., 2012. Evaluation of methyl chavicol concentration by different evaporation processes using central composite experimental design. *Separation and Purification Technology* 98, 464–471. doi:10.1016/j.seppur.2012.08.009.
- Mashima, J., Kodama, Y., Fujisawa, T., Katayama, T., Okuda, Y., Kaminuma, E., Ogasawara, O., Okubo, K., Nakamura, Y., Takagi, T., 2017. DNA Data Bank of Japan. *Nucleic Acids Research* 45, D25–D31. doi:10.1093/nar/gkw1001.
- Materi, W., Wishart, D.S., 2007. Computational systems biology in drug discovery and development: methods and applications. *Drug Discovery Today*. doi:10.1016/j.drudis.2007.02.013.
- McCall, C., 2014. Tibetans in Nepal struggle to access health care. *The Lancet*. doi:10.1016/S0140-6736(14)61712-8.
- McInerney, P., Adams, P., Hadi, M.Z., 2014. Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Molecular Biology International* 2014, 1–8. doi:10.1155/2014/287430.
- Medeiros, P.M.D., Soldati, G.T., Alencar, N.L., Vandebroek, I., Pieroni, A., Hanazaki, N., De Albuquerque, U.P., 2012. The use of medicinal plants by migrant people: Adaptation, maintenance, and replacement. *Evidence-based Complementary and Alternative Medicine*. doi:10.1155/2012/807452.
- Meier, Rudolf, Guanyang Zhang, and Farhan Ali. 2008. “The Use of Mean Instead of Smallest Interspecific Distances Exaggerates the Size of the ‘Barcoding Gap’ and Leads to Misidentification.” *Systematic Biology*. doi:10.1080/10635150802406343.
- McConkey, B.J., Sobolev, V., Edelman, M., 2002. The performance of current methods in ligand-protein docking. *Current Science* 83, 845–856. doi:10.1093/brain/awr286.
- Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D., Koornneef, M., 1998. *Arabidopsis thaliana*: A model plant for genome analysis. *Science*. doi:10.1126/science.282.5389.662.
- Meyerowitz, 1987. *Arabidopsis thaliana*. *Annual Review of Genetics* 21, 93–111. doi:10.1111/j.1399-3054.2009.01220.x.
- Michael, T.P., VanBuren, R., 2015. Progress, challenges and the future of crop genomes. *Current Opinion in Plant Biology*. doi:10.1016/j.pbi.2015.02.002.
- Miller JA, Miller EC, Phillips DH. 1982. The metabolic activation and carcinogenicity of alkenylbenzenes that occur naturally in many spices. In: Stich HF, editor. *Carcinogens and Mutagens in the Environment*, Vol. 1. Boca Raton, FL: CRC. pp 83–96.
- Ministry of Health and Family Welfare, 2003. The Drugs and Cosmetics Act, 1940 and Rules, 1945. *Archives of Neurology* 60, 191–196. doi:10.1001/archneur.60.4.643.

- Miraj, S., Kiani, S., 2016. A scientific correlation between dystemprament in Unani medicine and diseases: a systematic review. *Electronic physician* 8, 3240–3247. doi:10.19082/3240.
- Mishra, D., Awasthi, A., Mishra, P., 2014. Phylogenetic evolution studies on different varieties of genus *Ocimum* with special reference to Rewa district of Madhya Pradesh. *Sci Secure J Biotech* Volume 3, Issue 2, Pages 188-197.
- Mishra, P., Kumar, A., Nagireddy, A., Mani, D.N., Shukla, A.K., Tiwari, R., Sundaresan, V., 2016. DNA barcoding: An efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnology Journal* 14, 8–21. doi:10.1111/pbi.12419.
- Molina-Cano, J.L., 1989. Plant morphology and development. Varietal taxonomy. *La Cebada* 25–64; 27.
- Möller M., 2000. How universal are universal rDNA primers? A cautionary note for plant systematists and phylogeneticists. *Edinburgh Journal of Botany*; 57:151–156.
- Monette, M., 2012. The medicine of the prophet. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. doi:10.1503/cmaj.109-4228.
- Monro, R., 1997. Yoga therapy. *Journal of Bodywork and Movement Therapies* 1, 215–218. doi:10.1016/S1360-8592(97)80047-2.
- Moss, J., Yuan, C.-S., 2006. Herbal medicines and perioperative care. *Anesthesiology* 105, 441–2. doi:10.1001/jama.286.2.208.
- Mukherjee, P.K., Wahile, A., 2006. Integrated approaches towards drug development from Ayurveda and other Indian system of medicines. *Journal of Ethnopharmacology*. doi:10.1016/j.jep.2005.09.024.
- Murphy, a, Zhou, J., Goldsbrough, P.B., Taiz, L., 1997. Purification and immunological identification of metallothioneins 1 and 2 from *Arabidopsis thaliana*. *Plant physiology* 113, 1293–1301. doi:10.1104/pp.113.4.1293.
- Murphy, G.E.P., Romanuk, T.N., 2014. A meta-analysis of declines in local species richness from human disturbances. *Ecology and Evolution* 4, 91–103. doi:10.1002/ece3.909
- Murzin A. G. , Brenner S. E. , Hubbard T. , Chothia C., 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*;247:536–540.
- Needleman, S.B., Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* 48, 443–453. doi:10.1016/0022-2836(70)90057-4.
- Ngo, L.T., Okogun, J.I., Folk, W.R., c, 2013. 21st Century natural product research and drug development and traditional medicines. *Natural Product Reports* 30, 584–592. doi:10.1039/c3np20120a.
- Nicolson, D.H., 2002. Stone, plant, or animal, in: *Taxon*. pp. 7–10. doi:10.2307/1554958

- Nirmal, A., Pal, C., Otimenyin, S., Thanda A., Mostafa, E., Kumar, K., Sukalyan, A., Rajarajan, A. Subhash & Mandal, C., 2013. Contribution of Herbal Products In Global Market. Available from: Elachouri Mostafa, Oct 12, 2017.
- Norn, S., Kruse, P.R., Kruse, E., 2005. History of opium poppy and morphine. *Dansk medicinhistorisk arbog* 33, 171–184.
- Notredame, C., Higgins, D. G., & Heringa, J., Notredame, C., Higgins, D.G., Heringa, J., 2000. T-coffee: a novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* 302, 205–217. doi:10.1006/jmbi.2000.4042.
- Nowak, B., Miczyński, K., Hudy, L., 2004. Sugar uptake and utilisation during adventitious bud differentiation on in vitro leaf explants of “Węgierka Zwykła” plum (*Prunus domestica*).
- Pandey, A.K., Singh, P., Tripathi, N.N., 2014. Chemistry and bioactivities of essential oils of some *Ocimum* species: an overview. *Asian Pacific Journal of Tropical Biomedicine* 4, 682–694. doi:10.12980/APJTB.4.2014C77
- Plant Cell, Tissue and Organ Culture* 76, 255–260.
doi:10.1023/B:TICU.00000009247.94819.02.
- Nutton, V., 2004. *Ancient Medicine, Ancient Medicine*. Routledge Taylor & Francis Group. doi:10.4324/9780203490914.
- Obeng-Ofori, D., Reichmuth, C., 1997. Bioactivity of eugenol, a major component of essential oil of *ocimum suave* (wild.) against four species of stored-product coleoptera. *International Journal of Pest Management* 43, 89–94. doi:10.1080/096708797229040.
- Ogawa, Y., Dansako, T., Yano, K., Sakurai, N., Suzuki, H., Aoki, K., Noji, M., Saito, K., Shibata, D., 2008. Efficient and high-throughput vector construction and *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* suspension-cultured cells for functional genomics. *Plant and Cell Physiology* 49, 242–250. doi:10.1093/pcp/pcm181.
- Ogunmenwo, K.O., 2003. Cotyledon morphology : an aid in identification of *Ipomoea* taxa (*Convolvulaceae*). *Feddes Repertorium* 114, 198–203. doi:10.1002/fedr.200390022
- Ong, Q., Nguyen, P., Thao, N.P., Le, L., 2016. Bioinformatics Approach in Plant Genomic Research. *Current Genomics* 17, 368–378. doi:10.2174/1389202917666160331202956.
- Orengo C.A., Jones D.T., 1994. Thornton J.M. Protein superfamilies and domain superfolds. *Nature*;372:631–634.
- Pagani I., Liolios K., Jansson J., Chen I-M., Smirnova T., Nosrat B., Markowitz V.M., Kyrpides N.C. Kyrpides, 2012. The Genomes On-Line Database (GOLD) v. 4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.*;40(D1):D571–D579.
- Pai, V.R., Hegde, S., 2013. Research Article Preliminary Phytochemical Screening of Members of Lamiaceae Family: *International Journal of Pharmaceutical Sciences Review and Research* 21, 131–137.

Pandey, A.K., Singh, P., Tripathi, N.N., 2014. Chemistry and bioactivities of essential oils of some *Ocimum* species: an overview. *Asian Pacific Journal of Tropical Biomedicine* 4, 682–694. doi:10.12980/APJTB.4.2014C77.

Paton A., Harley R. M., Harley M.M., Hiltunen R., Holm Y. (Eds.), 1999. *Ocimum: An Overview of Classification and Relationship*, Harwood Academic Publishers, Amsterdam.

Pattanayak P, Behera P, Das D, Panda S, 2010. *Ocimum sanctum* Linn. A reservoir plant for therapeutic applications: An overview. *Pharmacogn Rev.* ;4:95–105.

Paul, D.B., 1988. The selection of the “Survival of the Fittest.” *Journal of the History of Biology* 21, 411–424. doi:10.1007/BF00144089.

Pazdera, H.J., 1966. Paper chromatography., in: *Encycl. Ind. Chem. Anal.* pp. 1–40.

Pesce, A.J., Gaver, R., Ashworth, R.B., 1974. *Chromatography.*, in: *Exp. Tech. Biochem.* Prentice-Hall, pp. 32–96.

Petsko, G. A. & Ringe, D., 2004. *Protein Structure and Function*, New Science Press.

Pichersky, E., and Gang, D.R., 2000. Genetics and biochemistry of secondary metabolites in plants: An evolutionary perspective. *Trends Plant Sci.* 5, 439–445.

Pieroni, A., Houlihan, L., Ansari, N., Hussain, B., Aslam, S., 2007. Medicinal perceptions of vegetables traditionally consumed by South-Asian migrants living in Bradford, Northern England. *Journal of Ethnopharmacology* 113, 100–110. doi:10.1016/j.jep.2007.05.009.

Pieroni, A., Torry, B., 2007. Does the taste matter? Taste and medicinal perceptions associated with five selected herbal drugs among three ethnic groups in West Yorkshire, Northern England. *Journal of Ethnobiology and Ethnomedicine* 3. doi:10.1186/1746-4269-3-21.

Premila, M., 2006. *Ayurvedic Botanicals – A Clinical Guide to the Healing Plants of Traditional Indian Medicine*. The Haworth Press, Inc., Binghamton, New York.

Prescott, A., Briddon, R., Harwood, W., 2010. Plant Transformation. *Transformation* 286, 1–41. doi:10.1002/9780470015902.a0003070.pub2.

Qiagen®, 2012. *DNeasy® Plant Handbook*. Qiagen®.

Raes, J., Bork, P., 2008. Molecular eco-systems biology: Towards an understanding of community function. *Nature Reviews Microbiology* 6, 693–699. doi:10.1038/nrmicro1935.

Rafatullah, S., Alqasoumi, S., 2008. Unani medicine: An integral part of health care system in Indian subcontinent. *European Journal of Integrative Medicine* 1, 39–40. doi:10.1016/j.eujim.2008.08.076.

Raina, A.P., Kumar, A., Dutta, M., 2013. Chemical characterization of aroma compounds in essential oil isolated from “Holy Basil” (*Ocimum tenuiflorum* L.) grown in India. *Genetic Resources and Crop Evolution* 60, 1727–1735. doi:10.1007/s10722-013-9981-4.

Raja, R.R., 2012. Medicinally Potential Plants of Labiatae (Lamiaceae) Family: An Overview. *Research Journal of Medicinal Plant*. doi:10.3923/rjmp.2012.203.

Ram, M., Jain, D.C., Mishra, H., Mandal, S., Abdin, M.Z., 2014. Recent advances to enhance yield of artemisinin: A novel antimalarial compound, in *Artemisia annua* L. plants, in: *Artemisia Annua - Pharmacology and Biotechnology*. Springer-Verlag Berlin Heidelberg, pp. 173–204. doi:10.1007/978-3-642-41027-7_11.

Rance, G. A., Marsh, D. H., Nicholas, R. J., & Khlobystov, A. N., 2010. UV-vis absorption spectroscopy of carbon nanotubes: Relationship between the π -electron plasmon and nanotube diameter. *Chemical Physics Letters*. doi:10.1016/j.cplett.2010.05.012.

Rao, R., 1987, *Encyclopedia of Indian Medicine* (2nd ed.), Dr. P.V. Parameshvara Charitable Trust, Bangalore, India.

Rastogi, S., Kalra, A., Gupta, V., Khan, F., Lal, R.K., Tripathi, A.K., Parameswaran, S., Gopalakrishnan, C., Ramaswamy, G., Shasany, A.K., 2015. Unravelling the genome of Holy basil: an “incomparable” “elixir of life” of traditional Indian medicine. *BMC genomics* 16, 413. doi:10.1186/s12864-015-1640-z.

Rastogi, S., Shukla, Y., Paul, B.N., Chowdhuri, D.K., Khanna, S.K., Das, M., 2007. Protective effect of *Ocimum sanctum* on 3-methylcholanthrene, 7,12-dimethylbenz(a)anthracene and aflatoxin B1 induced skin tumorigenesis in mice. *Toxicology and Applied Pharmacology* 224, 228–240. doi:10.1016/j.taap.2007.05.020.

Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., and Jackson, R. B., 2011. DNA tools and biotechnology. In *Campbell biology* (10th ed., pp. 408-435). San Francisco, CA: Pearson.

RefSeq, 2015. RefSeq Growth Statistics. URL <http://www.ncbi.nlm.nih.gov/refseq/statistics/>.
Reich E, Schibli A. Stationary Phases for Planar Separations - Plates for Modern TLC. *LC GC*. 2005;23:58–69.

Reid, G.M.G., 2009. Carolus Linnaeus (1707-1778): His life, philosophy and science and its relationship to modern biology and medicine, in: *Taxon*. pp. 18–31.

Renu, I. K., Haque, I., Kumar, M., Poddar, R., Bandopadhyay, R., Rai, A., & Mukhopadhyay, K. (2014). Characterization and functional analysis of eugenol O-methyltransferase gene reveal metabolite shifts, chemotype specific differential expression and developmental regulation in *Ocimum tenuiflorum* L. *Molecular Biology Reports*, 41(3), 1857–1870. <https://doi.org/10.1007/s11033-014-3035-7>.

Reshma, K., Kamalaksh, S., Bindu, Y., Pramod, K., Asfa, A., Amritha, D., Aswathi, S., Chandrashekar, R., 2012. Tulasi(*Ocimum Sanctum*) as radioprotector in head and neck cancer patients undergoing radiation therapy. *Biomedicine / [publiee pour l’A.A.I.C.I.G.]*.

Rishton, G.M., 2008. Natural Products as a Robust Source of New Drugs and Drug Leads: Past Successes and Present Day Issues. *American Journal of Cardiology* 101. doi:10.1016/j.amjcard.2008.02.007.

Roberson E., 2008. Medicinal plants at risk. Nature’s pharmacy, our treasure chest: why we must conserve our natural heritage. http://www.biologicaldiversity.org/publications/papers/Medicinal_Plants_042008_lores.pdf. Accessed (May, 2018).

Roberti di Sarsina, P., Ottaviani, L., Mella, J., 2011. Tibetan medicine: A unique heritage of person-centered medicine. *EPMA Journal*. doi:10.1007/s13167-011-0130-x.

Rocco, C.J., Dennison, K.L., Klenchin, V.A.V. a, Rayment, I., Escalante-Semerena, J.C., 2008. Construction and use of new cloning vectors for the rapid isolation of recombinant proteins from *Escherichia coli*. *Plasmid* 59, 231–7. doi:10.1016/j.bbi.2008.05.010.

Ronse De Craene, L.P., 2010. Floral diagrams: An aid to understanding flower morphology and evolution, *Floral Diagrams: An Aid to Understanding Flower Morphology and Evolution*. Cambridge University Press. doi:10.1017/CBO9780511806711.

Ronzón-Ortega, M., Hernández-Vergara, M.P., Pérez-Rostro, C.I., 2012. Hydroponic and aquaponic production of sweet basil (*ocimum basilicum*) and giant river prawn (*macrobrachium rosenbergii*). *Produccion hidropónica y acuapónica de albahaca (ocimum basilicum) y langostino malayo (macrobrachium rosenbergii)* 15, S63–S71.

Rosen, L., French, A., Sullivan, G., 2015. Complementary, Holistic, and Integrative Medicine: Yoga. *Pediatrics in Review* 36, 468–474. doi:10.1542/pir.36-10-468.

Ryder, V.L., 1954. On the morphology of leaves. *The Botanical Review* 20, 263–276. doi:10.1007/BF02872372

Sahoo N, Manchikanti P, Dey S. 2010. Herbal drugs: standards and regulation. *Fitoterapia* 81: 462–471.

Sahoo, N., Manchikanti, P., 2013. Herbal drug regulation and commercialization: an Indian industry perspective. *Journal of alternative and complementary medicine (New York, N.Y.)* 19, 957–63. doi:10.1089/acm.2012.0275.

Sai, 2010, *MEDICINAL PLANTS IN INDIA*, Blogspot, viewed 05/12/105, <<http://medicinalpltsinindiamp.blogspot.co.uk/>>.

Sakai, H., Lee, S.S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., Wakimoto, H., Yang, C.C., Iwamoto, M., Abe, T., Yamada, Y., Muto, A., Inokuchi, H., Ikemura, T., Matsumoto, T., Sasaki, T., Itoh, T., 2013. Rice annotation project database (RAP-DB): An integrative and interactive database for rice genomics. *Plant and Cell Physiology* 54. doi:10.1093/pcp/pcs183.

Saluta M., Bell P. A., 1998. Troubleshooting GST fusion protein expression in *E.coli*. *Amersham Pharmacia Biotech* 1–3.

Samuel, G., 2013. Introduction: Medicine and Healing in Tibetan Societies. *East Asian Science, Technology and Society* 7, 335–351. doi:10.1215/18752160-2333224.

Samuelson, J.C., 2011. Heterologous Gene Expression in *E. coli*. *Methods in Molecular Biology* 705, 195–209. doi:10.1007/978-1-61737-967-3.

Sandhu, D.S., Heinrich, M., 2005. The use of health foods, spices and other botanicals in the sikh community in London. *Phytotherapy Research* 19, 633–642. doi:10.1002/ptr.1714.

Sangster, G., 2014. The application of species criteria in avian taxonomy and its implications for the debate over species concepts. *Biological Reviews* 89, 199–214. doi:10.1111/brv.12051

Sasidharan, S., Chen, Y., Saravanan, D., Sundram, K.M., Yoga Latha, L., 2011. Extraction, isolation and characterization of bioactive compounds from plants' extracts. *African Journal of Traditional, Complementary and Alternative Medicines* 8, 1–10. doi:10.4314/ajtcam.v8i1.60483.

Satyamoorthy, K., Valiathan, M.V., Anchan, S., Bellampalli, R., Bhale, S., Bharadwaj, R., Bhat, B., Dedge, A., Dhumal, V., Gangadharan, G., Girijakumari, T., Rotti, H., Raval, R., Gopinath, P., Govindaraj, P., Halder, S., Joshi, K., Kabekkodu, S., Kamath, A., Kondaiah, P., Kukreja, H., Kumar, K.L.R., Nair, S., Nair, S.N.V., Nayak, J., Prasanna, B., Rashmishree, M., Sharanprasad, K., Thangaraj, K., Patwardhan, B., 2014. Determinants of Prakriti, the Human Constitution Types of Indian Traditional Medicine and its Correlation with Contemporary Science. *Journal of Ayurveda and Integrative Medicine* 5, 166. doi:10.4103/0975-9476.140478.

Savolainen, V., Cuénoud, P., Spichiger, R., Martinez, M.D.P., Crèvecoeur, M., Manen, J.F., 1995. The use of herbarium specimens in DNA phylogenetics: Evaluation and improvement. *Plant Systematics and Evolution* 197, 87–98. doi:10.1007/BF00984634.

Schinle T., Merfort I., 2012. New HPTLC methods for the fingerprint-characterization of *Ocimum sanctum* and *Achillea millefolium*. Master Thesis. Albert-Ludwigs-Universität Freiburg.

Schmid, R., Harris, J.G., Harris, M.W., 1995. Plant Identification Terminology: An Illustrated Glossary. *Taxon*. doi:10.2307/1222694.

Schmidt, B., Ribnicky, D.M., Poulev, A., Logendra, S., Cefalu, W.T., Raskin, I., 2008. A natural history of botanical therapeutics. *Metabolism: Clinical and Experimental* 57. doi:10.1016/j.metabol.2008.03.001.

Seberg, O., Petersen, G., 2009. How many loci does it take to DNA barcode a crocus? *PLoS ONE* 4. doi:10.1371/journal.pone.0004598.

Senanayake, U.M., Wills, R.B.H., and Lee, T.H., 1977. Biosynthesis of eugenol and cinnamic aldehyde in *Cinnamomum zeylanicum*. *Phytochemistry* 16, 2032–2033.

Sentausa, E., Fournier, P.E., 2013. Advantages and limitations of genomics in prokaryotic taxonomy. *Clinical Microbiology and Infection*. doi:10.1111/1469-0691.12181

Seth SD, Sharma B. 2004. Medicinal plants of India. *Indian J Med Res* 120:9–11.

Sgamma, T. 2012. Thesis. doi: <http://wrap.warwick.ac.uk/48973/>.

Sgamma, T., Lockie-Williams, C., Kreuzer, M., Williams, S., Scheyhing, U., Koch, E., Slater, A., Howard, C., 2017. DNA Barcoding for Industrial Quality Assurance. *Planta Medica*. doi:10.1055/s-0043-113448.

- Shah, U., Shah, R., Acharya, S., Acharya, N., 2013. Novel anticancer agents from plant sources. *Chinese Journal of Natural Medicines* 11, 16–23. doi:10.1016/S1875-5364(13)60002-3.
- Shaikh BT, Hatcher J., 2005. Complementary and alternative medicine in Pakistan: prospects and limitations. *eCAM*; 2:139-142.
- Shaw, J., Lickey, E.B., Schilling, E.E., Small, R.L., 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The Tortoise and the hare III. *American Journal of Botany* 94, 275–288. doi:10.3732/ajb.94.3.275.
- Shishodia S, Majumdar S, Banerjee S, Aggarwal B, 2003. Urosolic acid inhibits nuclear factor-kappaB activation induced by carcinogenic agents through suppression of IkappaBalpha kinase and p65 phosphorylation: Correlation with down-regulation of cyclooxygenase 2, matrix metalloproteinase 9, and cyclin D1. *Cancer Res*; 63:4375–83.
- Siddique, Y., Ara, G., Beg, T., Afzal, M., 2007. Anti-genotoxic effect of *Ocimum sanctum* L. extract against cyproterone acetate induced genotoxic damage in cultured mammalian cells. *Acta Biologica Hungarica* 58, 397–409. doi:10.1556/ABiol.58.2007.4.7.
- Siegel, J., Saukko, P., 2012. *Encyclopedia of forensic sciences*. Academic Press.
- Silva, M.R.R., Oliveira, J.G., Fernandes, O.F.L., Passos, X.S., Costa, C.R., Souza, L.K.H., Lemos, J.A., Paula, J.R., 2005. Antifungal activity of *Ocimum gratissimum* towards dermatophytes. *Mycoses* 48, 172–175. doi:10.1111/j.1439-0507.2005.01100.x.
- Silvertown, J., Dodd, M., Gowing, D., Lawson, C., McConway, K., 2006. Phylogeny and the hierarchical organization of plant diversity. *Ecology* 87. doi:10.1890/0012-9658(2006)87[39:PATHOO]2.0.CO;2.
- Simmonds, P., 2012. SSE: A nucleotide and amino acid sequence analysis platform. *BMC Research Notes* 5. doi:10.1186/1756-0500-5-50.
- Simon J.E., Quinn J., Murray R.G., Janik J. (Eds.), 1990. *Basil: A Source of Essential Oils*, Timber Press, Portland, Ore, USA.
- Simon, J., Morales, M., 1999. Basil: a source of aroma compounds and a popular culinary and ornamental herb. *Perspectives on new crops and new uses*; 499–505.
- Simpson, M.G., 2010. Diversity and Classification of Flowering Plants: Eudicots, in: *Plant Systematics*. Elsevier, pp. 275–448. doi:10.1016/B978-0-12-374380-0.50008-7
- Singh N, Hoette Y, Miller R., 2002. Tulsi ‘The Mother Medicine of Nature’ 1st Edn, International Institute of Herbal Medicine, Lucknow, pp. 29-42.
- Singh H., 2006. Prospects and challenges for harnessing opportunities in medicinal plants sector in India. *LEAD J*;2: 198–211.
- Singh N, Pandey MK, Sharma A, 2014. Indian medicinal plants: for hair care and cosmetics. *World J Pharm Sci* 2:1552–1556.

Singh, M., Kaur, R., Rajput, R., Mathur, G., 2017. Evaluating the therapeutic efficiency and drug targeting ability of alkaloids present in *Rauwolfia serpentina*. *International Journal of Green Pharmacy* 11, 132–142.

Singh, P., Kalunke, R.M., Giri, A.P., 2015. Towards comprehension of complex chemical evolution and diversification of terpene and phenylpropanoid pathways in *Ocimum* species. *RSC Adv.* 5, 106886–106904. doi:10.1039/C5RA16637C.

Singh, P.K., Baxi, D., Banerjee, S., Ramachandran, A.V., 2012. Therapy with methanolic extract of *Pterocarpus marsupium* Roxb and *Ocimum sanctum* Linn reverses dyslipidemia and oxidative stress in alloxan induced type I diabetic rat model. *Experimental and Toxicologic Pathology* 64, 441–448. doi:10.1016/j.etp.2010.10.011.

Sir Williams Monier, A, 2007. *Sanskrit English Dictionary*, 3rd Edition, New Delhi, Bharatiya Granth Niketan.

Skoog, F., Miller, C.O., 1957. Chemical regulation of growth and organ formation in plant tissues cultured in vitro. *Symp Soc Exp Biol* 11, 118–131. doi:10.1093/pcp/pcp176.

Smith, J.P.J., 2017. The herbarium. *Botanical Studies* 36, 1–7.

Smith, R.L., Adams, T.B., Doull, J., Feron, V.J., Goodman, J.I., Marnett, L.J., Portoghese, P.S., Waddell, W.J., Wagner, B.M., Rogers, A.E., Caldwell, J., Sipes, I.G., 2002. Safety assessment of allylalkoxybenzene derivatives used as flavouring substances - Methyl eugenol and estragole. *Food and Chemical Toxicology*. doi:10.1016/S0278-6915(02)00012-1.

Soultanas, P., 2011. The replication-transcription conflict. *Transcription* 2, 140–144. doi:10.4161/trns.2.3.15908.

Srivastava, M., 2011. High-performance thin-layer chromatography (HPTLC), High-Performance Thin-Layer Chromatography (HPTLC). Springer Berlin Heidelberg. doi:10.1007/978-3-642-14025-9.

Staats, M., Cuenca, A., Richardson, J.E., Ginkel, R.V. van, Petersen, G., Seberg, O., Bakker, F.T., 2011. DNA damage in plant herbarium tissue. *PLoS ONE* 6. doi:10.1371/journal.pone.0028448.

Storfer, A., Murphy, M.A., Evans, J.S., Goldberg, C.S., Robinson, S., Spear, S.F., Dezzani, R., Delmelle, E., Vierling, L., Waits, L.P., 2007. Putting the “landscape” in landscape genetics. *Heredity*. doi:10.1038/sj.hdy.6800917

Suanarunsawat T and Songsak T., 2005. Anti-hyperglycaemic and anti-dyslipidaemic effect of dietary supplement of white *Ocimum Sanctum* Linnean before and after STZ-induced diabetes mellitus. *Int J Diabetes Metab*; 13:18–23.

Suanarunsawat, T., Devakul Na Ayutthaya, W., Songsak, T., Thirawarapan, S., & Pongshompoo, S. (2009). Antioxidant Activity and Lipid-Lowering Effect of Essential Oils Extracted from *Ocimum sanctum* L. Leaves in Rats Fed with a High Cholesterol Diet. *Journal of Clinical Biochemistry and Nutrition*, 46(1), 52–59. <https://doi.org/10.3164/jcbrn.09-52>.

Subbarayappa, B.V., 1997. Siddha medicine: An overview. *Lancet*. doi:10.1016/S0140-6736(97)04223-2.

- Subramanian, A.R., Kaufmann, M., Morgenstern, B., 2008. DIALIGN-TX: Greedy and progressive approaches for segment-based multiple sequence alignment. *Algorithms for Molecular Biology* 3. doi:10.1186/1748-7188-3-6.
- Subramanian, S.V., Madhavan, V.R. (eds), 1984. *Heritage of the Tamils Siddha Medicine*, Madras, International Institute of Tamil Studies.
- Sudberg, S., Sudberg, E. M., Terrazas, J., Sudberg, S., Patel, K., Pineda, J., & Fine, B., 2010. Fingerprint analysis and the application of HPTLC to the determination of identity and quality of botanicals, from an industry perspective. *Journal of AOAC International*.
- Sulaiman, S.F., Sajak, A.A.B., Ooi, K.L., Supriatno, Seow, E.M., 2011. Effect of solvents in extracting polyphenols and antioxidants of selected raw vegetables. *Journal of Food Composition and Analysis* 24, 506–515. doi:10.1016/j.jfca.2011.01.020.
- Swarbreck D., Wilks C., Lamesch P., Berardini T.Z. Garcia- Hernandez, M.; Foerster, H.; Li, D.; Meyer, T.; Muller, R.; Ploetz, L., 2008. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.*;36(Suppl. 1):D1009–D1014.
- Swartz, M. (2010). HPLC detectors: A brief review. *Journal of Liquid Chromatography and Related Technologies*. <https://doi.org/10.1080/10826076.2010.484356>.
- Sweedler, J.V., 2002. The continued evolution of hyphenated instruments. *Analytical and Bioanalytical Chemistry* 373, 321–322. doi:10.1007/s00216-002-1358-z.
- Swenson, N.G., 2011. The role of evolutionary processes in producing biodiversity patterns, and the interrelationships between taxonomic, functional and phylogenetic biodiversity. *American Journal of Botany*, 98, 472–480.
- Tait, J., Mittra, J., 2004. Industry Challenges: Complexity and Innovation in the Pharmaceutical Industry. *Chemistry and Industry* 6, 1–4.
- Taiz, L., Zeiger, E., 2006. Secondary metabolites and plant defense, in: *Plant Physiology*. pp. 283–308.
- Tamokou, J. de D., Mbaveng, A.T., Kuete, V., 2017. Antimicrobial Activities of African Medicinal Spices and Vegetables, in: *Medicinal Spices and Vegetables from Africa: Therapeutic Potential Against Metabolic, Inflammatory, Infectious and Systemic Diseases*. Elsevier Inc., pp. 207–237. doi:10.1016/B978-0-12-809286-6.00008-X.
- Taylor, P.M., 1983. Plant and Animal Nomenclature in the Tobelorese Language. *Buletin LEKNAS [Special Issue, Halmahera dan Raja Ampat Sebagai Kesatuan Majemuk: Studi-Studi Terhadap Suatu Daerah Transisi]* 2, 41–76.
- Telles S., 2015. Ayurveda: an ancient science of holistic healing. DOI: [https://doi.org/10.5712/rbmfc7\(1\)510](https://doi.org/10.5712/rbmfc7(1)510).
- The C. elegans Sequencing Consortium, 1998. Genome Sequence of the Nematode *Caenorhabditis elegans*: A Platform for Investigating Biology. *Science* 282, 2012–2018. doi:10.1126/science.282.5396.2012.

The Plantlist, 2010. Version 1.1 [<http://www.theplantlist.org/browse/A/Lamiaceae/Ocimum/>]. Online. Accessed on 22/01/2018.

The Spectator, 2012. Ugandan Asians are part of Britain's secret weapon for success. [Online] Available from: <http://blogs.spectator.co.uk/2012/12/ugandan-asians-are-part-of-britains-secret-weapon-for-success/> [Accessed on 06/02/16].

Thomas M Walter, Rubia G, Sathiya E, 2009. A review of Ethics in Traditional Siddha. Siddha papers 4, 1–10.

Thornton, S.A., 1992. Use of differential agar media for detection of cloned DNA fragments in the tetracycline and chloramphenicol resistance genes of pBR322. Plasmid 27, 177–180. doi:10.1016/0147-619X(92)90019-7.

Töppner, K., Hansen, D., Herbig, E., 2014. HPLC Analysis: The Role of Ultrapure Water. HPLC Analysis 1–3.

Torii, K.U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R.F., Komeda, Y., 1996. The Arabidopsis ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. The Plant Cell 8, 735–746. doi:10.1105/tpc.8.4.735.

Evans, W. C., Evans, D., & Trease, G. E. (2009). Trease and Evans pharmacognosy. Edinburgh: Saunders/Elsevier.

Trower, M.K., 1996. A rapid PCR-based colony screening protocol for cloned inserts. Methods in molecular biology (Clifton, N.J.). doi:10.1385/0-89603-402-X:329.

U.S. Food and Drug Administration, 1994. Dietary Supplement Health and Education Act of 1994 [WWW Document]. US Department of Health and Human Services.

Udupa, K.N., 1975. THE AYURVEDIC SYSTEM OF MEDICINE IN INDIA. Newell, Kenneth W. (Ed.). Health by the People. Xii+206p. Illus. Maps. World Health Organization: Geneva, Switzerland. Isbn 92-4-156042-8 53–69.

Ueda-Nakamura, T., Mendonça-Filho, R.R., Morgado-Díaz, J.A., Korehisa Maza, P., Prado Dias Filho, B., Aparício Garcia Cortez, D., Alviano, D.S., Rosa, M. do S.S., Lopes, A.H.C.S., Alviano, C.S., Nakamura, C.V., 2006. Antileishmanial activity of Eugenol-rich essential oil from *Ocimum gratissimum*. Parasitology International 55, 99–105. doi:10.1016/j.parint.2005.10.006.

Uozumi, N., Kim, E.J., Rubio, F., Yamaguchi, T., Muto, S., Tsuboi, a, Bakker, E.P., Nakamura, T., Schroeder, J.I., 2000. The Arabidopsis HKT1 gene homolog mediates inward Na(+) currents in xenopus laevis oocytes and Na(+) uptake in *Saccharomyces cerevisiae*. Plant physiology 122, 1249–1259. doi:10.1104/pp.122.4.1249.

Uskoković, M.R., Grethe, G., 1973. The Cinchona Alkaloids. Alkaloids: Chemistry and Physiology 14, 181–223. doi:10.1016/S1876-0813(08)60222-5.

Vaidya ADB, Devasagayam TPA., 2007. Current status of herbal drugs in India: an overview. J Clin Biochem Nutr;4:1–11 [PMC free article] [PubMed].

Vassão, D.G., Gang, D.R., Koeduka, T., Jackson, B., Pichersky, E., Davin, L.B., Lewis, N.G., 2006. Chavicol formation in sweet basil (*Ocimum basilicum*): cleavage of an esterified C9 hydroxyl group with NAD(P)H-dependent reduction. *Org. Biomol. Chem.* 4, 2733–2744. doi:10.1039/B605407B.

Vayssières, M.P., Plant, R.E., Allen-Diaz, B.H., 2000. Classification trees: An alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science* 11, 679–694.

Venkatachalam, P., Kalaiarasi, K., Sreeramanan, S., 2015. Influence of plant growth regulators (PGRs) and various additives on in vitro plant propagation of *Bambusa arundinacea* (Retz.) Wild: A recalcitrant bamboo species. *Journal of Genetic Engineering and Biotechnology* 13, 193–200. doi:10.1016/j.jgeb.2015.09.006.

Vieira, R.F., Grayer, R.J., Paton, A., Simon, J.E., 2001. Genetic diversity of *Ocimum gratissimum* L. based on volatile oil constituents, flavonoids and RAPD markers. *Biochemical Systematics and Ecology* 29, 287–304. doi:10.1016/S0305-1978(00)00062-4.

Vijay Kumari, 2012. Simplified extraction of good quality genomic DNA from a variety of plant materials. *African Journal of Biotechnology* 11. doi:10.5897/AJB11.2366.

Villas-Bôas, S.G., Roessner, U., Hansen, M.A.E., Smedsgaard, J., Nielsen, J., 2006. Metabolome Analysis: An Introduction, *Metabolome Analysis: An Introduction*. doi:10.1002/9780470105511.

Viña, A., Murillo, E., 2003. Essential Oil Composition from Twelve Varieties of Basil (*Ocimum* spp) Grown in Colombia. *Journal of the Brazilian Chemical Society* 14, 744–749. doi:10.1590/S0103-50532003000500008.

Voeks, R.A., 1990. Sacred Leaves of Brazilian Candomblé. *Geographical Review* 80:118–131.

Volpato, G., Godínez, D., Beyra, A., Barreto, A., 2009. Uses of medicinal plants by Haitian immigrants and their descendants in the Province of Camagüey, Cuba. *Journal of Ethnobiology and Ethnomedicine* 5. doi:10.1186/1746-4269-5-16.

Vongsak, B., Sithisarn, P., Mangmool, S., Thongpraditchote, S., Wongkrajang, Y., Gritsanapan, W., 2013. Maximizing total phenolics, total flavonoids contents and antioxidant activity of *Moringa oleifera* leaf extract by the appropriate extraction method. *Industrial Crops and Products* 44, 566–571. doi:10.1016/j.indcrop.2012.09.021.

Vyas, P., Mukhopadhyay, K., 2014. Development of a rapid and high frequency *Agrobacterium rhizogenes* mediated transformation protocol for *Ocimum tenuiflorum*. *Biologia (Poland)* 69, 765–770. doi:10.2478/s11756-014-0375-7.

Wachtel-Galor, S., Benzie, I., 2011. Herbal Medicine: An introduction to its history, usage, regulation, current trends, and research needs, in: *Herbal Medicine: Biomolecular and Clinical Aspects*. CRC Press, Taylor & Francis, Boca Raton, FL, pp. 1–11.

Wagner H, Ulrich-Merzenich G., 2009. Synergy research: approaching a new generation of phytopharmaceuticals. *Phytomedicine*; 16:97–110. doi: 10.1016/j.phymed.2008.12.018.

- Wakdikar S., 2004. Global health care challenge: Indian experiences and new prescriptions. *Electron J Biotechnol* ;7:214–220.
- Waldhoff, D., Parolin, P., 2011. Morphology and Anatomy of Leaves, in: Junk, W.J., Piedade, M.T.F., Wittmann, F., Schöngart, J., Parolin, P. (Eds.), *Amazonian Floodplain Forests*. Springer Netherlands, pp. 179–202. doi:10.1007/978-90-481-8725-6_9.
- Wang J, Pichersky E, 1999. Identification of specific residue involved in substrate discrimination in two plant O-methyltransferases. *Arch Biochem Biophys* 368:172–180. doi:10.1006/abbi1999.1304
- Wang, J., 1999. Molecular Characterization of an O-Methyltransferase Involved in Floral Scent Production in *Clarkia breweri*. Ph.D. Dissertation. (Ann Arbor, MI: University of Michigan).
- Wang, J., Pichersky, E., 1998. Characterization of S-adenosyl-L-methionine:(Iso)eugenol O-methyltransferase involved in floral scent production in *Clarkia breweri*. *Archives of Biochemistry and Biophysics* 349, 153–160. doi:10.1006/abbi.1997.0452.
- Wang, M., Zhu, X., Wang, K., Lu, C., Luo, M., Shan, T., Zhang, Z., 2018. A wheat caffeic acid 3-O-methyltransferase TaCOMT-3D positively contributes to both resistance to sharp eyespot disease and stem mechanical strength. *Scientific Reports* 8. doi:10.1038/s41598-018-24884-0.
- Warrier P K., 1995. In: *Indian Medicinal Plants*. Longman O, editor. New Delhi: CBS publication; p. 168.
- Warrier, M., 2011. Modern Ayurveda in Transnational Context. *Religion Compass* 5, 80–93. doi:10.1111/j.1749-8171.2011.00264.x.
- Wei, A.; Shibamoto, T. Antioxidant/lipoxygenase inhibitory activities and chemical compositions of selected essential oils. *J. Agr. Food Chem.* 2010, 58, 7218-7225.
- WHO traditional medicine strategy 2014-2023.
- WHO, 2000. General guidelines for methodologies on research and evaluation of traditional medicine, World Health Organization Geneva.
- WHO, 2002. Monographs on selected medicinal plants. *Folium Ocimi Sancti*. 2:206–216.
- WHO, 2003. Traditional medicine. [Online] Media centres. Available from: <http://www.who.int/mediacentre/factsheets/2003/fs134/en/> [Accessed 05/04/2016].
- WHO, 2011. The World Medicines Situation 2011. Medicines prices, availability and affordability, The World Medicines Situation.
- Wiesemüller, B., Rothe, H., 2006. Interpretation of bootstrap values in phylogenetic analysis. *Anthropologischer Anzeiger; Bericht über die biologisch-anthropologische Literatur* 64, 161–5.

Wink, M. Introduction: biochemistry, physiology and ecological functions of secondary metabolites, in: M. Wink (Ed.), *Biochemistry of Plant Secondary Metabolites*, Annu Plant Rev, second ed. Wiley-Blackwell, Hoboken, New Jersey 2010, pp. 1–19.

Wink, M., 2000. Interference of alkaloids with neuroreceptors and ion channels. *Studies in Natural Products Chemistry* 21, 3–122. doi:10.1016/S1572-5995(00)80004-6.

Wink, M., 2015. Modes of Action of Herbal Medicines and Plant Secondary Metabolites. *Medicines*, 2(3), 251–286. <https://doi.org/10.3390/medicines2030251>.

Wiseman RW, Miller EC, Miller JA, Liem A. 1987. Structure–activity studies of the hepatocarcinogenicities of alkenylbenzene derivatives related to estragole and safrole on administration to pre-weanling male C57BL/6J3C3H/HeJ F1 mice. *Cancer Res* 47:2275–2283.

Withgott, J., 2000. Is it `So Long, Linnaeus'? *BioScience*. doi:10.1641/0006-3568(2000)050[0646:IISLL]2.0.CO;2

World Health Organization (WHO), 2005. National Policy on Traditional Medicine and Regulation of Herbal Medicines. Geneva: Report of WHO global survey.

World Health Organization, 2003. WHO guidelines on good agricultural and collection practices (GACP) for medicinal plants. *World Health* 99, 67–73. doi:10.1017/CBO9781107415324.004.

WORLD HEALTH ORGANIZATION, W. H. O. (2010) Traditional Medicine. Fact sheet, 134.

Xie J. , Schultz P. G., 2005. An expanding genetic code. *Methods*;36:226–238.

Xie, Z., Kapteyn, J., Gang, D.R., 2008. A systems biology investigation of the MEP/terpenoid and shikimate/phenylpropanoid pathways points to multiple levels of metabolic control in sweet basil glandular trichomes. *Plant Journal* 54, 349–361. doi:10.1111/j.1365-313X.2008.03429.x.

Xie, Z., Zhang, Z., Cao, Z. *et al.* *Appl Microbiol Biotechnol*, 2017. An external substrate-free blue/white screening system in *Escherichia coli*. 101: 3811. <https://doi.org/10.1007/s00253-017-8252-2>.

Yauk, Y.K., Chagné, D., Tomes, S., Matich, A.J., Wang, M.Y., Chen, X., Maddumage, R., Hunt, M.B., Rowan, D.D., Atkinson, R.G., 2015. The O-methyltransferase gene MdoOMT1 is required for biosynthesis of methylated phenylpropenes in ripe apple fruit. *Plant Journal* 82, 937–950. doi:10.1111/tpj.12861.

Yesilada, E., 2011. Contribution of traditional medicine in the healthcare system of the Middle East. *Chinese Journal of Integrative Medicine* 17, 95–98. doi:10.1007/s11655-011-0651-0.

Yildiran, S.T., Yanikoglu, B., Abdullah, E., 2014. Plant identification using local invariants, in: *Signal Processing and Communications Applications Conference (SIU)*, 2014 22nd. pp. 2094–2097. doi:10.1109/SIU.2014.6830674.

Yip, P.Y., Chau, C.F., Mak, C.Y., Kwan, H.S., 2007. DNA methods for identification of Chinese medicinal materials. *Chinese Medicine*. doi:10.1186/1749-8546-2-9.

Zambryski, P., Holsters, M., Kruger, K., Depicker, A., Schell, J., Van Montagu, M., Goodman, H.M., 1980. Tumor DNA structure in plant cells transformed by *A. tumefaciens*. *Science* 209, 1385–1391. doi:10.1126/science.6251546.

Zaya, D.N., Ashley, M.V., 2012. Plant Genetics for Forensic Applications, in: Sucher, N.J., Hennell, J.R., Carles, M.C. (Eds.), *Plant DNA Fingerprinting and Barcoding*. pp. 35–52. doi:10.1007/978-1-61779-609-8_4.

Zhang Y , Baranov P.V. , Atkins J.F. , Gladyshev V.N., 2005. Selenocysteine and pyrrolysine use dissimilar decoding strategies. *J. Biol. Chem.*;280:20740–20751.

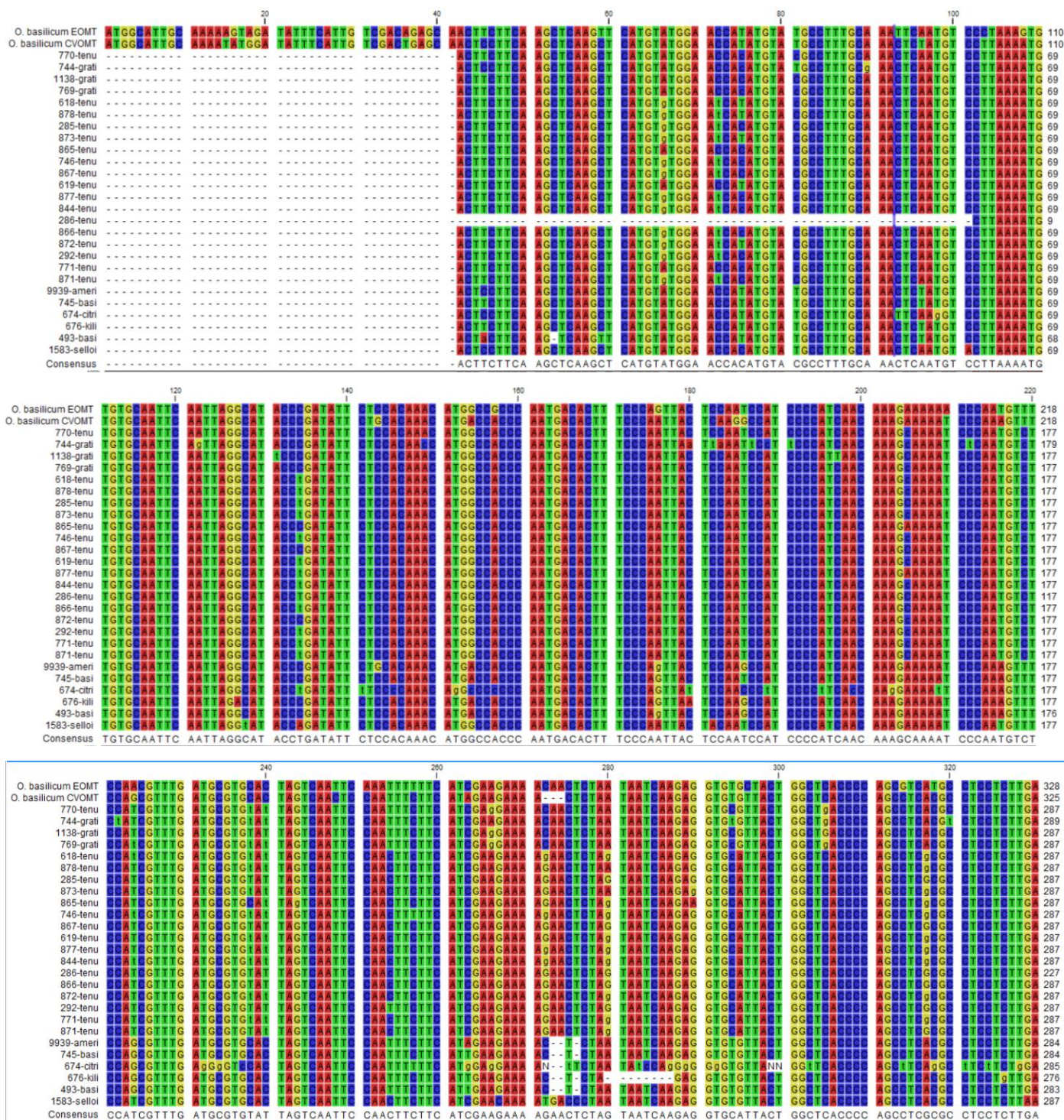
Zhang, X., Henriques, R., Lin, S.S., Niu, Q.W., Chua, N.H., 2006. Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nature Protocols* 1, 641–646. doi:10.1038/nprot.2006.97.

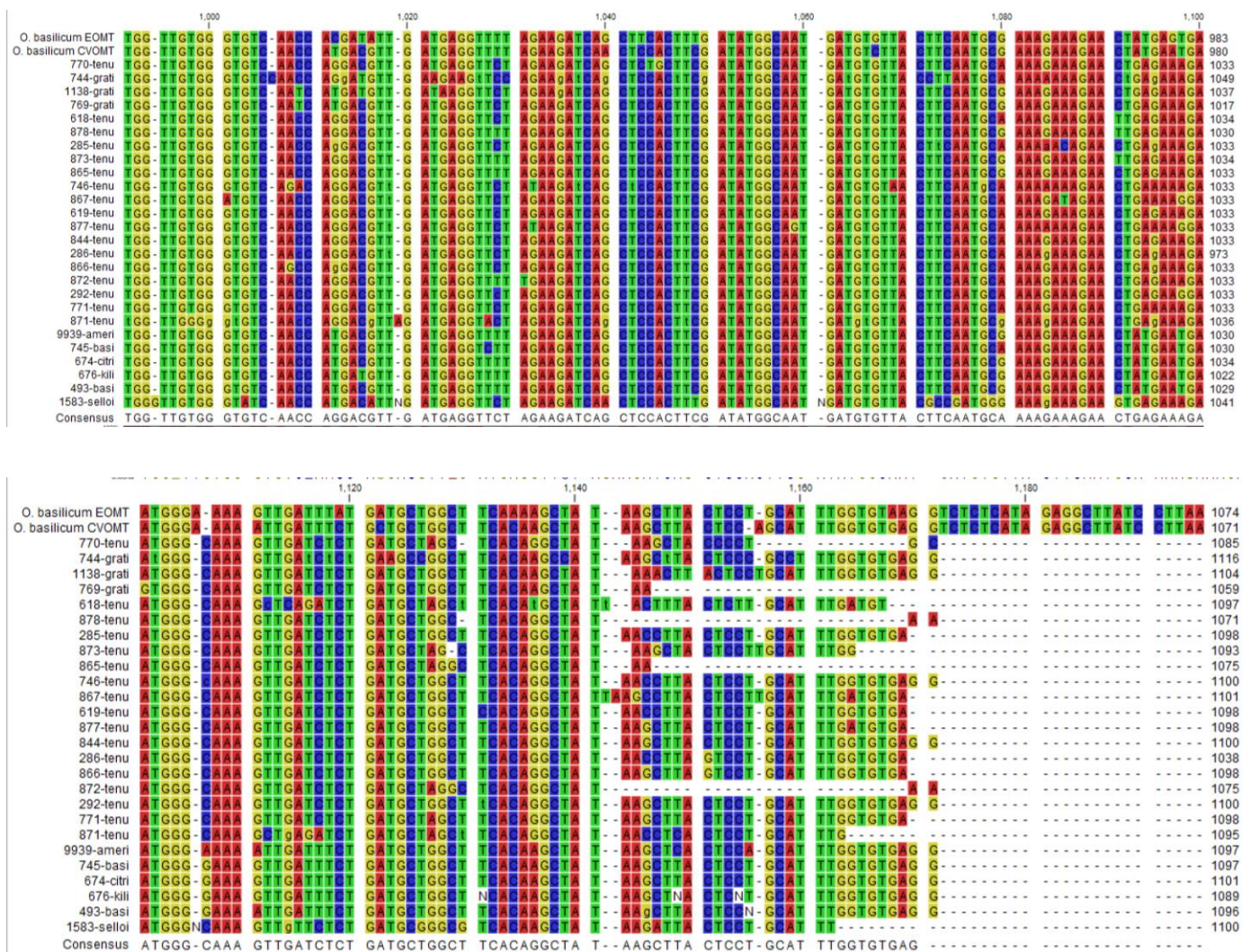
Zhu, Y.P., Woerdenbag, H.J., 1995. Traditional Chinese herbal medicine. *Pharmacy World & Science* 17, 103–112.

Zvi, M.M.B., Shklarman, E., Masci, T., Kalev, H., Debener, T., Shafir, S., Ovadis, M., Vainstein, A., 2012. PAP1 transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers. *New Phytologist* 195, 335–345. doi:10.1111/j.1469-8137.2012.04161.x.

Appendix

Appendix 1 Full nucleotide alignment with reference sequences





Appendix 2 Different size introns splicing from genomic DNAs

1. *O. tenuiflorum* - 618 (91 BP)

ACTTCTTCAAGCTCAAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATGTGCAA
TTCATTAGGCATACCTGATATTCTCCACAAACATGGCCACCAATGACACTTTCCCAATTACTCCAATCCATCC
CCATCAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTAAATTCCAACTTTTTCATCGAAG
AAAAGAACTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCAGCCTCGCGCCTCCTCTTGAAAGGGGCGCC
TTTGACTGTGGCACCCTTGTCCAAGTCATTTTGATCCCCTTTACAAACCCATGGCACCATATGAGCGAAT
GGTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAA
ATGAGCCGAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCAGATGTACTCAC
TAAGGACTACAAGCATGTGATTAGAGGGGAATTTAGAACGTTGGTTGATGTAGGGGGTGGTGACGGAACGA
TGGCCAAAGCAATTCGTTGAAACAGTGCCACCATAAAATGCACAGTTATTGACCTCCACATGTCACTGGCT
GGCTTGAAAGCAACGATAGCTTGAGCTATATTGGAGGAGACAAGTCCGGTCTATCCCTCCGCCGAGCCA
GTTCTTCTCATG**GTAATTCCTTAAGCCCCATCCACGCTAGCTATTTATATCATCTTATATTGAAAACAAAAC**
ACTAATGATGAGGGGAAAATACTGACCAGTCTATAATACAGACTGGAACGATGAGGAGGGCCTCAAAGTCTT
GAAGAGATGAAAAGATGCCGTGGTATTGGAGGGAAGGTGATTATAATCGATGTGGTTGTGGGTGTCAAAC
AGGACGTTGATGAGGTTGTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAAGA

ATGAACTGAGAAAGAATGGGCAAAGTTCAGATCTGATGCTCGCTTCATATGCTATTACTTTACTCTTGCAATTTG
ATGT

2. *O. tenuiflorum* - 619 (91 BP)

ACTTCTTCAAGCTCAAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATGTGCAA
TTCAATTAGGCATACCTGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTTACTCCAATCCAT
CCCCATCAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCAACTTCTTCATCGA
AGAAAAGAAGCTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCAGCCTCGCGCTCCTCTTGAAAGGGGCG
CCTTTGACAGTGGCACCCCTTGCCAAGTCATTTTGGATCCCACTTTTCAAAACCCATGGCACCATATGAGCGA
ATGGTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCA
AATGAGCCGAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACATGTACTCA
CTAAGGACTACAAGCATGTGATTGGGGGAATTAGAACGTTGGTTGATGTAGGGGGTGGTGACGGAACGATG
GCTAAAGCAATCGTTGAAGCAGTGGCCACCATAAAATGCACAGTTATTGACCTCCACATGTCGTGGCTGGCT
TGGAAGCACCAGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGCAGTTCT
TCTCAAGGTAAATTCCTTAAGCCCCATCCCACACAAGCTATTATATATTATCATATATTGAAAACAAAACATTAA
TGATGAGGGGAAAATACTGAACAGTCTATAATACAGGACTGGAACGATGAGGAGGGCCTCAAAATCTTGAAGA
GATGCAAAGATGCCGTCGGTATTGGAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTCAACCAGGACG
TTGATGAGGTTTTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAAAGAAAGAAC
TGAGAAAGAATGGGCAAAGTTGATCTCTGATGCTGGCTTACAGGCTATAACCTTACTCCTGCATTTGGTGTG
A

3. *O. citriodorum* - 674 (93 BP)

CCTTGGCAAATTTAAGGCCCTTAAAAGGTCCAATCCAATTGGGCATCCCGGATATTTTCCCAACCAGGCCCC
CCAATGCCATTTTCCCGGTTATTCCAACCTTCCCTTACCAAGGAAAATTCCCAAAGTTTCCAGCGTTGGAG
GGGTCCATTAGTCAATTCCAATTTTTCCATGGAGGAAAATTTAATATCCAGGGGGGGGGTTATGGGTTCCCC
CCACCTTCAGGCTTCTTTTGGAGGGGGGCCCTTTGGATGGGGCCCCCTGTTCAAGGTGGTTTGGGTTCCCA
TTTTCAAAACCCAGGCCATATAAGGGGGGAAGGGTTTAAACATGAGAACCCGCCCCCAGTTTGGGGGCG
CCAAAGGGAGACACGTTTGGGGGGAGGTTGCCAAATAACGCCAGCCAGGGGTGGATTTTGAAGACGCTT
GGGGTGGGACGCAAGGCTTGGGGCACATGTACTCCCCAGGACTACAAGCATGTGATTGATGGAATAAGAA
CATTGGTCGATGTTGGGGGTGTAATGGAACGATGGCTAAAGCTATCGTAGAAGCAATGCCACCATGAAAT
GCACAGTTCTTGACCTACCACATGTTGTAAGTGGGTTAGAAAGCACCGACAGATTGAGCTATATTGGGGGAAA
CATGTTCCAGTCTATCCCCTCTGCAGATAAAATCTTCTAAAGGTAAATTACTTAAAGCCCCATTAACTTAAAT
TATTATATATAATTATATATTGAAAAAAAATAATATAGTGATGATTAATACTACACAGTCTATAATACAG
AGGGGGACCAATAAGGAGGGCATCAATATCGTGAAGAGATGTAAAGATGCAGTTGGTGTGGAGGGAAGGT
GATAATCATCGATGTGGTTGTGGGTGTCAACCATGACGTTGATGAGGTTTTAGAAGATCAGCTCCACTTCGAT
ATGGCAATGATGTGTTACTTCAATGCGAAAGAAAGAACTATGAATGAATGGGGAAAGTTGATTTCTGATGCT
GGCTTACAAGCTATAAGCTTACTCCTGCATTTGGTGTGAGG

4. *O. kilimandscharicum* - 676 (90 bp)

TTCTTCAAGCTCAAGCTCATGTAGGGAACCATATGTATGCCTTTGCAAACCTCTATGTCCTTAAAATGTGCAATTC
AATTAGACATACCCGATATTCTCCACAAACATGACCACCCAATGACACTTTCCAGTTAATCCAAGCCATCCCCA
TCAACAAAGAAAAATCCCAAAGTTTCCAGCGTTTGATGCGTGCCTAGTCAATTCCAATTTCTTCATTGAAGAA
AACTCTAGAGGTGTGTTACTGGCTCACCCAGCCTCACGCCTCCTGTTGAAGGGGGCGCCTCTAACTGTGGCA
CCCCCTGTCCAAGTCGTTTTGGATCCCACTTTACAAGCCCATGGCATTATATGAGTGAATGGTTTAAACATGA
GAACCACGCCACCCAGTTTGAAGCAGCAAATGGATGCACGTTTTGGGAGAAGTTAGCAAATAAGCCAAGCAT
GGGTAGATTTTTTATGAAGCTATGAGTTGTGACTCAAGGCTTGTGGCACATGTACTACCAAGGACTACAAG
CATGTGATTGATGGAATAAGAACATTGGTCGATGTTGGGGGTGGTAATGGAACGATGGCTAAAGCTATCGTC
GAAGCAGTGGCCACCATGAAATGCATAGTTATTGACCTACCACATGTTGTGGCTGGGTTGGAAAGCAATGACA

GATTAAGCTATATTGGGGGAGCCATGTTCCAGTCTATCCCTTCTGCAGATGCGATTCTTCTCAAGGTAAATTAC
TTAAAGCCCCATTAAACTTGAATTATTATATATCATTATACATTGAAAACAAAATAATAATGATGAGTAAAATA
CTAGCAGTCTATAATACACGATTGGGACGATGAGGAAGGCCTCAAAATCTTGAAGAGATGTAAAGATGCAGT
TGGTATTGGAGGGAAGGTGATAATCATCGATGTGGTTGTGGGTGTCAACCATGATGTTGATGAGGTTTTAGA
AGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCCAAAGAAAGAACTATGAATGAATGGG
AAAGTTGATTCTGATGCTGGCTTCACAAGCTATAAGCTTACTCCTGCATTGTTGGTGTGAGG

5. *O. gratissimum* - 744 (103 bp)

CACATGTAGCGCCTTGGCGAATCTCAATCTCCTTAAAATGTGCAATTGAGTTGGCCTTTCCGCATATTTCCACA
AACAGGCCACCCAAGGACATTTTCCCATTACTCCAATTCCTTCCCATCAACAAAGCAAATCTCAATGTCTCTA
TCGTTTGAGGCGTGATTAGTCAATTCCAACCTTCTTCATCGAGGAAAACAACTCTAATAATCAAGAGGTGCGTT
ACGGGCTGACCCAGCCTCACGTCTCCTTTGAAGGGGGCGCCTTTGACTGTGGCACCCCTTGTCCAAGTCATT
TTGGATCCCACTTTCACAAACCCATGGCACCATATGAGTGAATGGCTTACACACGAGCACCACGCCACACAGT
TTGAGCCAGAAAAGGGATGCATGTTTTGGGAGAAGTTAACAAATGAGCCAAGCATGAGTAGATTTTTTGAGG
AAGCTATGAGTTGTGACTCGAGGCTCGTGGCACATGTACTACCAAGGACTACAAGCATGTTTTTCAGGGAAT
CAAAACATTGGTTGATGTAGGGGGTGGTAACGGAACGATGGCTAGAGCTATTGTGGAAGCAGTGCCACCAT
GAAATGCACAGTTATTGACCTCCACATGTTGTGGCTGGCTGGAAAAGCACCGATACCTTGAGCTATATGGGG
GGAGACATGCTCCGATCTATCCCCTCGCCCGATGCAGTTCTTCTCAAGGTAAATTCCTTTAAGCTTAAGCCCC
GTCATCAAATTAGAAAAATATTATATATCCTCATGCCTTGAAAACAAAACAATAATGATGAGCAAAAAAATG
AACAGTCTATAAAACACGACCGGAACCAATAAGGAGGGCCTCAAAATCCTGAAAAGATGCAAAGATGCCGTC
CGGATTGGAAGGAAAGGGATTATTATCCACCTGGTTGTGGGTGTCCAACCAGAATGTTGAAGAGGGTCCAGA
AAAACAACTCCACCTCCATATGGCAATGAAGTGGTACCTTAATGCAAAAAAAGAACCAGAAAAAGAAAGGGC
AAAGTTGAACGCCGAAGCCGGCTTCACAAGCCATAAGCCTACTCCCGCCTTGGTGTGAGG

6. *O. basilicum* - 745 (91 bp)

ACTTCTCAAGCTCAAGCTCATGTATGGAACCATATAGTATGCCTTTGCAAACCTCTATGTCCTTAAAATGTGCAA
TTCAATTAGGCATACCCGATATTCTCCACAAACATGACCACCCCAATGACACTTTCCAGTTACTCCAATCCATC
CCCATCAACAAAGAAAAATCCCAAAGTTTCCAGCGTTTGATGCGTGCACTAGTCAATTCCAATTTCTTCATTGA
AGAAAACCTAATAATCAAGAGGTGTGTTACTGGCTCACCCAGCCTCACGCCCTCCTTGAAGGGGGCGCCT
CTGACTGTGGCACCCCTTGTCCAAGTCGTTTTGGATCCCACTTTCACAAACCCATGGCATTATATGAGTGAATG
GTTTAAACATGAGAACCACGCCACCCAGTTTGAGGCAGCAAATGGATGCACGTTTTGGGAGAAGTTAGCAAA
TAAGCCAAGCATGGGTAGATTTTTTGATGAAGCTATGAGTTGTGACTCAAGGCTTGTGGCACATGTACTCACC
AAGGACTACAAGCATGTGATTGATGGAATAAGAACATTGGTCGATGTTGGGGGTGGTAATGGATCGATGGCT
AAAGCTATCGTCGAAGCAGTGCCACCATGAAATGCACAGTTCTTGACCTACCACATGTTGTAGCTGGGTTGG
AAAGCACCGACAGATTAAGCTATATTGGGGGAGACATGTTCCAGTCTATCCCTTCTGCAGATGCAATTCTTCTC
AAGGTAAATTACTTAAAGCCCCATTAAACTTCAATTATTATACATAATTATACATTGAAAACAAAATAATAATG
ATGAGTAAAATACTAAACAGTCTATAATACACGATTGGGACGATGAGGAGGGCATCAAAATCTTGAAGAGAT
GTAAAGATGCAGTTGGTATTGGAGGGAAGGTGATAAATCGATGTGGTTGTGGGTGTCAACCATGACGTTG
ATGAGGTCTTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAAGAAAGAACTAT
GAATGAATGGGGAAAGTTGATTCTGATGCTGGCTTCACAAGCTATAAGCTTACTCCTGCATTGTTGGTGTGAGG

7. *O. tenuiflorum* - 746 (91 bp)

TTCTTCAAGCTCAAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATGTGCAATTC
AATTAGGCATACCTGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCCCCA
TCAACAAAGCAAATCCCAATGTCTCCATCGTTTGATGCGTGATTAGTCAATTCCAACCTTTTTCATCGAAGAA
AAGAACTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCAGCCTCGCGCCTCCTTGTAAAGGGGGCGCCTT
TGACAGTGGCACCCCTTGTCCAAGTCATTTTGGATCCCACTTTCACAAACCCATGGCACCATATGAGCGAATGG
TTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAAATG

AGCCGAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACATGTACTCACTAA
GGACTACAAGCATGTGATTGAGGGGAATTAGAACGTTGGTTGATGTAGGGGGTGGTGACGGAACGATGGCT
AAAGCATCGTTGAAACAGTGCCCAACATAAAATGCACAGTTATTGACCTCCACATGTCTGGCTGGCTTGA
AAGCACCGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGCAGTTCTTCTCA
AGGTAAATTCCTTAAGCCCCATCCCACTGAAGCTATTATATATCATCATATATTGAAAAACAAACATTAATGAT
GAGGGAAAATACTGAACAGTCTATAATACAGGACTGGAAGATGAGGAGGGCCTCAAAATCTAGAAGAGAT
GAAAAGATGACGTAGGTATGGGAGGGAAGGTGATTATAATAGAAGTGGTTGTGGGTGTCAGACAGGACGTA
GATGAGGTTCTATAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATACAAAAAAGAAGT
AGAAAGAATGGGCAAAGTTGATCTCTGATGCTGGCTTACAGGCTATAACCTTACTCCTGCATTTGGTGTGAG
G

8. *O. gratissimum* - 769 (74 bp)

ACTTCTTCAAGCTCAAGCTCATGTATGGAACCAcATGTAcGCCTTTGCAAACCTCAATGTCCTTAAATGTGCAAT
TCAATTAGGCATACCCGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCCC
CATCAACAAAGcAAAAATCCCAATGTcTCCATcGTTTGATGCGTGTAtTAGTCAATTCCAATTTCTTCATCGAgGAA
AACAACCTAATAATCAAGAGGTGcGTTACTGGCTgACCCAGCCTCACGCCTCCTCTTGAAGGGGGCGCCTCT
GACTGTGGCACCCCTTGTTCAAGTcATTTTGGATCCCACTTTCAAAACCCATGGCACCATATGAGcGAATGGc
TTACACATGAGAACCACGCCACACAGTTTGAGGCAGCAAATGGATGCACGTTTTGGGAGAAGTTAGCAAATG
AGCCAAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGACTCGAGGCTCGTGGCACATGTACTACCAA
GGACTACAAGCATGTGATTGAGGGAATCAAAACATTGGTTGATGTAGGGGGTGGTAATGGAACGATGGCT
AgAGCTATCGTTGAAGCAGTGCCCAACATGAAATGCACAGTTATTGACCTCCACATGTcGTGGCTGGCTTGG
AAAGCACTGATAgCTTgAGCTATATTGGGGGAGACATGTTCCgGTCTATCCCTTCTGCcGATGCAGTTCTTCTCA
AGGTAAATTCCTTAAGCTATTATATATCTTATAATTATAAAACAAAACCAATAATGATGAGCAAAAACATGAAC
AGTCTATAATACACGACTGGAACGATGAGGAAGGCCTCAAAATCTTGAATAGATGCAAAGATGCGGTGCGTA
TTGGAGGTAAGGTGATTATTATCGACGTGGTTGTGGGTGTCAATCATGACGTTGATGAGGTTCTAGAAGATCA
GCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCGAAAGAAAGAACTGAGAAAGAGTGGGCAAAGTT
GATCTCTGATGCTGGCTTACAAGCTATAA

9. *O. tenuiflorum* - 770 (91 bp)

ACTTCTTCAAGCTCAAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAATGTGCAA
TTCAATTAGGCATACCTGATATTCTCCACAAACATGGCCACCCAATGAAACTTTCTCAATTACTCCAATCCATC
CCCATCAACAAAGCAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAATTCTTCATCGAA
GAAAAGAACTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCAGCCTAGCGCCTCCTCTTGAAAGGGGGCGC
CTCTGACAGTGGCACCCCTTGTTCAAGTcATTTTGGATCCCACTTTCAAAACCCATGGCACCATATGAGCGAA
TGGTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAA
ATGAGCCGAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACATGTACTCAC
TAAGGACTACAAGCATGTGATTGGGGGAATTAGAACGTTGGTTGATGTAGGGGGTGGTGACGGAACGATGG
CTAAAGCAATCGTTGAAGCAGTGCCCAACATAAAATGCACAGTTATTGACCTCCACATGTCTGGCTGGCTT
GGAAAGCACCGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGCAGTTCTT
CTCAAGGTAAATTCCTTAAGCCCCATCCCAACAAAGCTATTATATATCATCATATATAGAAAAACAAACAGTAA
TGATGAGGGAAAATACTGAACAGTCTATAATACACGACTGAAACGATGAGGAGGGCCTCAAAATCTTGAAGA
GATGAAAAGATGCCGTGCGTATTGAAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTCAGCCAGGACG
TTGATGAGGTTCTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAGAAAGAAC
TGAGAAAGAATGGGCAAAGTTGATCTCTGATGCTGGCTTACAGGCTATAACCTTACTCCTGCATTTGGTGTG
A

10. *O. tenuiflorum* - 771 (91 bp)

ACTTCTTCAAGCTCAAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAATGTGCAA
TTCAATTAGGCATACCTGATATTCTCCACAAACATGGCCACCCAATGAAACTTTCTCAATTACTCCAATCCATC

CCCATCAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAACCTCTTCATCGAA
GAAAAGAACTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCAGCCTAGCGCCTCCTCTTGAAAGGGGCGC
CTCTGACAGTGGCACCCCTTGTTCAAGTCATTTTGATCCCACTTTACAAACCCATGGCACCATATGAGCGAA
TGGTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAA
ATGAGCCGAGCATGGGTAGATTTTTTGATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACATGTACTCAC
TAAGGACTACAAGCATGTGATTGGGGGAATTAGAAGCTTGTTGATGTAGGGGGTGGTGACGGAACGATGG
CTAAAGCAATCGTTGAAGCAGTGCCACCATAAAATGCACAGTTATTGACCTCCACATGTCTGGCTGGCTT
GGAAAGCACCGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGCAGTTCTT
CTCAAGGTAAATTCCTTAAGCCCCATCCCACACAAGCTATTATATATCATCATATATAGAAAACAAAACAGTAA
TGATGAGGGAAAATACTGAACAGTCTATAATACACGACTGAAACGATGAGGAGGGCCTCAAAATCTTGAAGA
GATGAAAAGATGCCGTCGGTATTGAAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTGAGCCAGGACG
TTGATGAGGTTCTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAAGAAAGAAC
TGAGAAAGATGGGCAAAGTTGATCTCTGATGCTGGCTTACAGGCTATAACCTTACTCCTGCATTTGGTGTG
A

11. *O. tenuiflorum* - 844 (91 bp)

ACTTCTTCAAGCTCAAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAATGTGCAA
TTCAATTAGGCATACCTGATATTCTCCACAAACATGGCCACCAATGACACTTTCCCAATTACTCCAATCCATCC
CCATCAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAACCTCTTCATCGAAG
AAAAGAACTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCAGCCTCGCGCCTCCTCTTGAAAGGGGACGCC
TTTGACAGTGGCACCCCTTGTTCAAGTCATTTTGATCCCACTTTACAAACCCATGGCACCATATGAGCGAAT
GGTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAA
ATGAGCCGAGCATGGGTAGATTTTTTGATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACATGTACTCAC
TAAGGACTACAAGCATGTGATTGGGGGAATTAGAAGCTTGTTGATGTAGGGGGTGGTGACGGAACGATGG
CTAAAGCAATCGTTGAAGCAGTGCCACCATAAAATGCACAGTTATTGACCTCCACATGTCTGGCTGGCTT
GGAAAGCACCGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGCAGTTCTT
CTCAAGGTAAATTCCTTAAGCCCCATCCCACACAAGCTATTATATATTATCATATATTGAAAACAAAACATTAAT
GATGAGGGAAAATACTGAACAGTCTATAATACACGACTGGAACGATGAGGAGGGCCTCAAAATCTTGAAGAG
ATGCAAAGATGCCGTCGGTATTGGAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTCAACCAGGACGT
TGATGAGGTTCTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAAGAAAGAAC
GAGAAAGAATGGGCAAAGTTGATCTCTGATGCTGGCTTACAGGCTATAAGCTTACTCCTGCATTTGGTGTGA
GG

12. *O. tenuiflorum* - 865 (91 bp)

ACTTCTTCAAGCTCAAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAATGTGCAA
TTCAATTAGGCATACCTGATATTCTCCACAAACATGGCCACCAATGACACTTTCCCAATTACTCCAATCCATCC
CCATCAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAACCTCTTCATCGAA
GAAAAGAACTCTAGTAATCAAGAAGTGCATTACTGGCTCACCCAGCCTCGCGCCTCCTCTTGAAAGGGGCGC
CTCTGACTGTGGCACCCCTTGTTCAAGTCATTTTGATCCCACTTTACAAACCCATGGCACCATATGAGCGAA
TGATTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAA
ATGAGCCGAGCATGGGTAGATTTTTTGATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACATGTACTCAC
TAAGGACTACAAGCATGTGATTGGGGGAATTAGAAGCTTGTTGATGTAGGGGGTGGTGACGGAACGATGG
CCAAAGCAATCATTGAAGCAGTGCCACCATAAAATGCACGGTTATTGACCTCCACATGTCTGGCTGGCTT
GGAAAGCACCGATAGCTTGAGCTATATTGGAGGAGACATGTTACGGTCTATCCCTTCTGCCGATGCAGTTCT
TCTCAAGGTAAATTCCTTAAGCCCCATCCCACCCAGCTATTTTATATTTTCTCATATTGAAAACAAAACATTAA
TGATGAGGGAAAACACTGAACAGTCTATAATACACGACTGGAACGATGAGGAGGGCCTCAAAATCTTGAAGA
GATGCAAAGATGCCGTCGGTATTGGAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTCAACCAGGACG
TTGATGAGGTTTTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCGAAAGAAAGAAC
TGAGAAAGAATGGGCAAAGTTGATCTCTGATGCTAGGCTCACAGGCTATAA

13. *O. tenuiflorum* - 867 (91 bp)

ACTTCTTCAAGCTCAAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATGTGCAA
TTCAATTAGGCATACCTGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCC
CCATCAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAATTCTTCATCGAAG
AAAAGAACTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCCAGCCTCGCGCCTCCTCTTGAAAGGGGCGCC
TTTGACAGTGGCACCCTTGTTCCAAGTCATTTTGGATCCCACTTTTCAAAACCCATGGCACCATATGAGCGAAT
GGTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAA
ATGAGCCGAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACATGTACTCAC
TAAGGACTACAAGCATGTGATTGGGGGAATTAGAAGTTGGTTGATGTAGGGGGTGGTGACGGAACGATGG
CTAAAGCAATCGTTGAAGCAGTGGCCACCATAAAATGCACAGTTATTGACCTCCACATGTCGTGGCTGGCTT
GGAAAGCACCAGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGCAGTTCTT
CTCAAGGTAAATTCCTTAAGCCCCATCCACAGAAGCTATTATATATTATCATATATTGAAAACAAAACATTAAT
GATGAGGGGAAAATACTGACCAGTCTATAATACACGACTGGAACGATGAGGAGGGCCTCAAAATCTTGAAGAG
ATGCTAAGATGCCGTCGGTATTGAAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTCAACCAGGACGT
TGATGAGGTTCTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAAGATAGAAGT
GAAAAGGAATGGGCAAAGTTGATCTCTGATGCTGGCTTCACAGGCTATTAAGCCTTACTCCTTGCAATTTGATGT
GA

14. *O. tenuiflorum* - 871 (93 bp)

ACTTCTTCAAGCTCAAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATGTGCAA
TTCAATTAGGCATACCTGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCC
CCATCAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAATTCTTCATCGAAG
AAAAGAACTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCCAGCCTCGCGCCTCCTCTTGAAAGGGGCGCC
TTTGACAGTGGCACCCTTGTTCCAAGTCATTTTGGATCCCACTTTTCAAAACCCATGGCACCATATGAGCGAAT
GGTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAA
ATGAGCCGAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACATGTACTCAC
TAAGGACTACAAGCATGTGATTGGGGGAATTAGAAGTTGGTTGATGTAGGGGGTGGTGACGGAACGATGG
CTAAAGCAATCGTTGAAGCAGTGGCCACCATAAAATGCATAGTTATTGACCTCCACATGTCGTGGCTGGCTT
GGAAAGCACCAGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGCAGTTCTT
CTCAAGGTAAATTCCTTAAGCCCCATCCCACTGCTAGCTATAATATATTATCTTATATTGAAAACAATACATTAA
ATGATGAGGGGAAAATACTGAACAGTCTATAATACACGACTGGAACGATGAGGAGGGCCTCAAAATCTTGAAG
AGATGCAAAGATGCCGTCGGTATTGGAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTCAACCAGGAC
GTTAGATGAGGTTATAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAAGATAGA
ACTGAGAAAGATGGGCAAAGCTCATCTCTGATGCTGGCTTCACAGGCTATAACCTCACTCCTGCATTTG

15. *O. tenuiflorum* - 872 (93 bp)

ACTTCTTCAAGCTCAAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATGTGCAA
TTCAATTAGGCATACCTGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCC
CCATCAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAATTCTTCATCGAAG
AAAAGAACTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCCAGCCTCGCGCCTCCTCTTGAAAGGGGCGCC
TTTGACTGTGGCACCCTTGTTCCAAGTCATTTTGGATCCCACTTTTCAAAACCCATGGCACCATATGAGCGAAT
GATTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAA
TGAGCCGAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACATGTACTCACT
AAGGACTACAAGCATGTGATTGGGGGAATTAGAAGCTTGGTTGATGTAGGGGGTGGTGACGGAACGATGGC
TAAAGCAATCCTTGAAGCAGTGGCCACCATAAAATGCACAGTTATTGACCTCCACATGTCATGGCTGGCTTG
GAAAGCACCAGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGCAGTTCTTCT
CAAGGTAAATTCCTTAAGCCCCATCCCACTCCAAGCTATTATATATCATCTTATATTGAAAACAAAACATGAAAT
GATGAGGGGAAAATACTGAACAGTCTATAATACACGACTGGAACGATGAGGAGGGCCTCAAAATCTTGAAGAG
ATGAAAAGATGCCGTCGGTATTGGAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTCAACCAGGACGT

TAGATGAGGTTCTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAAGATTGAAC
TGAGAAAGAATGGGCAAAGCTGATCTCTGATGCTGGCTTACAGGCTATAACCTCACTCCTGCATTGG

16. *O. tenuiflorum* - 873 (92 bp)

ACTTCTTCAAGCTCAAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATGTGCAA
TTCAATTAGGCATACCTGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCC
CCATCAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAACCTTCTCATCGAAG
AAAAGAAGCTCTAATAATCAAGAAGTGCATTACTGGCTCACCCCAGCCTCGCGCCTCCTCTTGAAAGGGGCGCC
TCTGACTGTGGCACCCTTGTCCAAGTCATTTTGGATCCCCTTTACAAACCCATGGCACCATATGAGCGAAT
GGTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAA
ATGAGCCGAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCAGCATGTACTCAC
TAAGGACTACAAGCATGTGATTGGGGGAATTAGAAGCTTGGTTGATGTAGGGGGTGGTGACGGAACGATGG
CCAAAGCAATTCTTGAAGCAGTGCCACCATAAAATGCACGGTTATTGACCTCCACATGTCATGGCTGGCTTG
GAAAGCACCGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGCAGTTCTTCT
CAAGGTAAATTCCTTAAGCCCCATCCACCCCAGCTATTATATATTTATCATATATTGAAAACAAAACATTAATG
ATAAGAAAAAACACTAAACAGTCTATAATACACGACTGGAACGATGAGGAGGGCCTCAAATCTTGAAGAGA
TGCAAAGATGCCGTGGTATTGGAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTCAACCAGGACGTT
GATGAGGTTTTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCGAAAGAAAGAATTG
AGAAAGAATGGGCAAAGTTGATCTCTGATGCTAGCTCACAGGCTATAAGCTACTCCTTGCATTGG

17. *O. tenuiflorum* - 878 (89 bp)

ACTTCTTCAAGCTCAAGCTCATGTGTGGAATCATATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATGTGCAA
TTCAATTAGGCATACCCGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCC
CCATCAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAACCTTCTCATCGAAG
AAAAGAAGCTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCCAGCCTCGCGCCTCCTCTTGAAAGGGGCGCC
TTTGACAGTGGCACCCTTGTCCAAGTCATTTTGGATCCCCTTTACAAACCCATGGCACCATATGAGCGAAT
GGTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAA
ATGAGCCGAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCAGCATGTACTCAC
TAAGGACTACAAGCATGTGATTGGGGGAATTAGAAGCTTGGTTGATGTAGGGGGTGGTGACGGAACGATGG
CTAAAGCAATGCATTGAAGCAGTGCCACCATAAAATGCACAGTTATTGACCTCCACATGTCGTGGCTGGCT
TGGAAGCACCGATAGCTTGAGCTATATTGGAGGAGACATGTTTCGGTCTATCCCTTCTGCCGATGCAGTTCTT
CTCAAGGTAAATTCCTTAAGCCCCATCCACACTAGCTATTATATATTATCATATTTTGAAGAAAAACATTAATGA
TAAGAAAAAGCACTATACTGTCTATAATACTACTGGAACGATGAGAAGGGCCTCAAATCTTGACGAGATG
AAAAGATGCCATCGGTATTGGAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTCAACCAGGACGTTGA
TGAGGTTTTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCGAAAGAAAGAATTGAG
AAAGAATGGGCAAAGTTGATCTCTGATGCTGGCTCACAGGCTATAA

18. *O. selloi* – 1583 (94 bp)

ACTCCTTCAAGCTCAAGCTCATGTATGGAACCACATGTATGCCTTTGCAAACCTCAATGTACTTAAAATGTGCAA
TTCAATTAGGtATACCAGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTACTACAATCCATCC
CCATCAACAAAGAAAAATCCCAATGTTTCCATCGTTTGATGCGTGCCTAGTCAATTCCAACCTTCTCATCGAAG
AAAATGACCTAATAATCAAGAGGTGTGTTACTGGCTCACCCCAGCCTCACGCCTCCTCTTAAAGGGAGAGCC
TCTAACTGTGGCACCCTTATCCAAGCCATTTACATCCCACATTACAAACCCATGGCACCATATGAGTGAAT
GGTTTAAAGATGAGAAACATGCCACACAGTTGAGGCAACAAATAtATCCACAGTTTGGGAGAcGTTAGCgAA
TGAGCCTAGCcTGAATAGcTTgTTTAATAAAGCTATGAGTTGTGACTCGgGGCTTATGGCtCATGTACTCgtCAA
GGACTAtAAGCATGTGATTGAGGGAATCAGAACATTGGTTCGATGTAGGGGGTGGCgATGGAACGATGGCTAA
AGCTATCGTTGAAGCAGTGCCACCATAAAATGCACTGTTCTTGACCTCCACATGTTGTGGATtGCTTGGAAA
GCActAATAgGTTgAGCTATATTGGGGGAGACATGTTCCAGTCTATCCCTTCTGcTATGCAATTCTTCTTAAGG
TAATTTCTTAAAGCCTTATTAAACTCAAACCTGTTATATATCATCAGTAGCCTTGAAAACAAAAATAATACTGATG

AGTAAAAACCCTGAACAGTGTATAATACATGATTGGAACGACAAGGAGAGCCTAAAAATCTTGAAGAGAtGC
AgAGATGCGGTCTGCACTGGAAGGGAAGGTGATTATCATTGGATGTGGGTGTGGGTATCAACCATGACATT
NGATGAGGTTCTAGAAGATCAACTCCACTTTGATATGGCAATNGATGTGTTACGCCGATGGGAAAgAAAGAA
GTGAGAAAGAATGGGNCAAAGTTgTTCTCTGATGCGGGCGTCACAAGCTATAAGATTACTCTGCATT

19. *O. gratissimum* – 1138 (95 bp)

ACTTCTTCAAGCTCAAGCTCATGTATGGAACCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATGTGCAA
TTCAATTAGGCATtCCCGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCC
CCATTAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAATTTCTTCATCGAgG
AAAACAACCTAATAATCAAGAGGTGCGTTACTGGCTGACCCCAGCCTCACGCCTCCTCTTGAAGGGGGCGCC
TCTGACTGTGGCACCCTTGTCCAAGTCATTTTGGATCCCACTTTACAAAACCCATGGCACCATATGAGCGAAT
GGCTTACACAcGAGCACCACACCACACAGTTTGAGGCAGCAAATGGATGCATGTTTTGGGAGAAGTTAGCAA
ATGAGCCGAGCATGAGTAGATTTTTTGTGAAGCTATGAGTTGTGACTCGAGGCTCGTgGCACATGTACTCAC
CAAGGACTACAAGCATGTGATTGAGGGAATCAAAACATTGTTGATGTAGGGGGTGGTAATGGAACGATGG
CTAgAGCTATCGTTGAAGCAGTGCCCAACATGAAATGCACAGTTATTGACCTCCACATGTCGTGGCTGGCTTG
GAAAGCACTGATAGCTTTGAGCTATATTGGGGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGCAGtTCTTCT
CAAGtTAAATTCCTTAAGCTTAACCCCGTCACATTAGAACTATTATATATCCTCATACCTTGAAAAACAAACATT
AATGATGAGCAAAACATGAACAGTCTATAATACACGACAGGAAAGATAAGGAAGGCCTCAAAATCTTGAAT
AGATACAAAGATgCAGTCGGTATTGGAGGTAAGGTGATTATTATCGACGTGGTTGTGGGTGTCAATCATGATG
TTGATAAGGTTCTAGAAgATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCGAAAGAAAGAAC
TGAGAAAGAATGGGCAAAGTTGATCTCTGATGCTGGCTTACAAGCTATAAACTTACTCTGCATTGTTGGTGTG
AGG

20. *O. tenuiflorum* – 877 (91 bp)

ACTTCTTCAAGCTCAAGCTCATGTgTGGAATcAcATGTAcGCCTTTGCAAACCTCAATGTCCTTAAAATGTGCAAT
TCAATTAGGCATACctGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCCC
CATCAACAAAGAAAAATCCCAATGTCTCCATCGTTTGATGCGTGTAtTAGTCAATTCCAATTCTTCATCGAAGA
AAAgAACTCTAgTAATCAAGAGGTGCaTTACTGGCTCACCCCAGCCTCgCGCCTCCTCTTGAAaGGGGCGCCTTT
GACaGTGGCACCCTTGTCCAAGTCATTTTGGATCCCACTTTACAAAACCATGGCACCATATGAGCGAATGGT
TtActCATGAGcACCATGCCACACAGTTTGAGGCGGCAAAATGGATGCATGTTTTGGGAaAAGTTAGCAAATGA
GCCgAGCATGGGTAGATTTTTTGTGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACATGTACTCACTAAG
GACTACAAGCATGTGATTGgGGGAATTAGAACGTTGGTTGATGTAGGGGGTGGTGACGGAACGATGGCTAA
AGCAATtGTTGAAGCAGTGCCCAACATGAAATGCACgGTTATTGACCTCCACATGTCGTGGCTGGCTTGAA
AGCACCGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGCAGtTCTTCTCAA
GGTAAATTCCTTAAGCCCCATCCACACAAGCGATAATATATTACCTCATATTGAAAAACAAACAGTAATGAAA
AGGAAAAATACTAAACAGTCTATAATACAggACTGgAACGATGAGGAGGGCCTCAAAATCTtGAAGAGATGCA
AAGATGCCGAAGGTATTGgAGGGAAGGtGATTATcAtGtGTTGTGGGTGTCAACCAGGACGtGATGA
GGTCTATAAGATCAGCTCCACTTCGATATGGCAGTGATGTGTTACTTCAATGCAAAAAAAGAAGTGAAG
GAATGGGCAAAGTTGATCTCTGATGCTGGCTTACAGGCTATAAGCTTACTCTGCATTGATGTGA

21. *O. tenuiflorum* – 866 (91 bp)

ACTTCTTCAAGCTCAAGCTCATGTgTGGAATcAcATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATGTGCAAT
TCAATTAGGCATACctGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCCC
CATCAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAAtTCTTCATCGAAGA
AAAGAACTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCCAGCCTCGCGCCTCCTCTTGAAAGGGGGCGCCT
TTGACAGTGGCACCCTTGTCCAAGTCATTTTGGATCCCACTTTACAAAACCATGGCACCATATGAGCGAATG
GTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAAATGGATGCATGTTTTGGGAAAAAGTTAGCAAAT
GAGCCGAGCATGGGTAGATTTTTTGTGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACATGTACTCACTA
AGGACTACAAGCATGTGATTGGGGGAATTAGAACGTTGGTTGATGTAGGGGGTGGTGACGGAACGATGGCT

AAAGCAATCGTTGAAGCAGTGCCCAACATAAAATGCACAGTTATTGACCTCCACATGTCGTGGCTGGCTTGG
AAAGCACCGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCAGATGCAGTTCTTCTC
AAGGTAAATTCCTTAAGCCCCATCACACAAGCTATTATATATTATCATATATTGAAAAACAAACATTAATGA
TGAGGGAAAATACTGAACAGTCTATAATACACGACTGGAACGATGAGGAGGGCCTTAAATCTAGAAGAGAT
GCaAAGATGCCGTCgGTATTGgAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTCAGCCAgGACGTTGA
TGAGGTTCTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAAGAAAGAACTGAg
AAAGAATGGGCAAAGTTGATCTCTGATGCTGGCTTCACAGGCTATAAGCTTAGTCCTGCATTTGGTGTGA

22. *O. basilicum* – 493 (91 bp)

ACTaCTTCAAGTCAAGTTCATGTATGGAACCATATGTATGCCTTTGCAAACCTCTATGTCCTTAAATGTGCAATT
CAATTAGGCATACCCGATATTCTCCACAAACATGACCACCAATGACACTTTCCCAgTTACTCCAAGCCATCCCC
ATCAACAAAGAAAAATCCCAAAGTTTCCAGCGTTTGATGCGTGCACTAGTCAATTCCAATTTCTTCATtGAAGA
AAACTCTAATAATCAAGAGGTGTGTTACTGGCTCACCCAGCCTCACGCCTCCTCTTGAAGGGGGCGCCTcTGA
CTGTGGCACCCCTTGTCaAGTCGTTTTGGATCCCACTTTCAAAACCCATGGCATTATATGAGTGAATGGTTT
AAACATGAGAACCACGCCACCCAGTTTGAGGCAGCAAATGGATGCACGTTTTGGGAGAAGTTAGCAAATAAG
CCAAGCATGGGTAGATTTTTTGATGAAGCTATGAGTTGTGACTCAAGGCTTGTCACATGTACTACTAAGG
ACTACAAGCATGTGATTGATGGAATAAGAACATTGGTCGATGTTGGGGGTGGTAATGGAACGATGGCTAAAG
CTATCGTCGAAGCAGTGCCCAACATGAAATGCACAGTTCTTGACCTACCACATGTTGTGGCTGGGTTGAAAG
CACCGACAAATTAAGCTATATTGGGGGAGACATGTTCCAGTCTATCCCTTCTGCAGATGCAATCTTCTCAAGG
TAAATTACTTAAAGCCCCATTAACTTGAATTATTATATATCATTATACATTGAAAAACAAATAATAATGATGAG
TAAATACTAAACAGTCTATAATACACGATTGGGACGATGAGGAGGGCCTCAAAATCTTGAAGAGATGTAA
GATGCAGTTGGTATTGGAGGGAAGGTGATAATCATCGATGTGGTTGTGGGTGTCAACCATGACGTTGATGAG
GTTTTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCGAAAGAAAGAACTATGAATG
AATGGGGAAAATTGATTTCTGATGCTGGCTTCACAAGCTATAAgCTTACTCCNGCATTTGGTGTGAGG

23. *O. tenuiflorum* – 292 (91 bp)

ACTTCTTCAAGCTCAAGCTCATGTgTGGAAtCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAATGTGCAAT
TCAATTAGGCATACcTGATATTCTCCACAAACATGGCCACCAATGACACTTTCCCAATTACTCCAATCCATCCC
CATCAACAAAGcAAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAATTTCTTCATCGAAGA
AAAGAACTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCAGCCTCGCGCCTCCTCTTGAAAGGGGGCGCCT
TTGACAGTGGCACCCCTTGTCaAGTCATTTTGATCCCACTTTCAAAACCCATGGCACCATATGAGCGAATG
GTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCAtTTTTGGGAAAAGTTAGCAAAT
GAGCCGAGCATGGGTAGATTTTTTGATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCAGATGTACTACTA
AGGACTACAAGCATGTGATTGGGGGAATTAGAAGCTTGTTGATGTAGGGGGTGGTGACGGAACGATGGCT
AAAGCAATCGTTGAAGCAGTGCCCAACATAAAATGCACAGTTATTGACCTCCACATGTCGTGGCTGGCTTGG
AAAGCACCGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGcGATGCAGTTCTTCTC
AAGGTAAATTCCTTAAGCCCCATCCACACAAGCTATTATATATTATCATATATTGAAAAACAAACATTAATGAT
GAGGGAAAATACTGAACAGTCTATAATACACGACTGGAACGATGAGGAGGGCCTCAAAATCTtGAAGAGATG
CAaAGATGCCGTCGGTATTGGAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTCAACCAGGACGTTGA
TGAGGTTcTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAAGAAAGAACTGAG
AAGGAATGGGCAAAGTTGATCTCTGATGCTGGCTtCACAGGCTATAAGCTTACTCCTGCATTTGGTGTGAGG

24. *O. tenuiflorum* – 285 (91 bp)

CTTAAATGTGCAATTCAATTAGGCATACCTGATATTCTCCACAAACATGGCCACCAATGACACTTTCCCAATT
ACTCCAATCCATCCCATCAACAAAGCAAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAA
cTTCTTCATCGAAGAAAAGAACTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCAGCCTCGCGCCTCCTCT
TGAAAGGGGGCGCTTTGACAGTGGCACCCCTTGTCaAGTCATTTTGATCCCACTTTCAAAACCCATGGCAC
CATATGAGCGAATGGTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCACGTTTTGGG

AAAAGTTAGCAAATGAGCCgAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGATGCGAGGCTCGTGG
CACATGTACTCACTAAGGACTACAAGCATGTGATTGGGGGAATTAGAACGTTGGTTGATGTAGGGGGTGGT
ACGGAACGATGGCTAAAGCAATCGTCGAAGCAGTGGCCACCATAAAATGCACAGTTATTGACCTCCCACATGT
CGTGGCTGGCTTGAAAGCACCGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCc
GATGCAGTTCTTCTCAAGGTAAATTCCTTAAGCCCCATCCACACAAGCTATTATATATTATCATATATAGAAAA
CAAAAAATTAATGAAGAGGGGAAAATACTGAACAGTCTATAATACACGACTGGAACGATGAGGAGGGCCTCAA
AATCTTGAAGAGATGcAAAGATGCCGTCGGTATTGgAGGGAAGGtGATTATCATCGATGTGGTTGTGGGTGTC
AAcCAGGACGTtGATGAGGTTCTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAA
AgAAAGAACTGAgAAAGAATGGGCAAAGTTGATCTCTGATGCTGGCTTCACAGGCTATAACCTTAGTCCTGCA
TTTGGTGTGA

25. *O. tenuiflorum* – 286 (91 bp)

ACTTCTTCAAGCTCAAGCTCATGTgTGGAAcCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATGTGCAAT
TCAATTAGGCATACcTGATATTCTCCACAAACATGGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCCC
CATCAACAAAGCAAATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAACCTTCTTCATCGAAG
AAAAGAActTAGTAATCAAGAGGTGCATTACTGGCTCACCCAGCCTCGCGCTCCTCTTGAAAGGGGCGCC
TTTGACAGTGGCACCCTTGTCCAAGTCATTTTGGATCCCACTTTCACAAACCCATGGCACCATATGAGCGAAT
GGTTTACTCATGAGCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAA
ATGAGCCGAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACATGTACTCAC
TAAGGACTACAAGCATGTGATTGGGGGAATTAGAACGTTGGTTGATGTAGGGGGTGGTGACGGAACGATGG
CTAAAGCAATCGTTGAAGCAGTGGCCACCATAAAATGCACAGTTATTGACCTCCCACATGTCGTGGCTGGCTT
GGAAAGCACCGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCcGATGCAGTTCTTC
TCAAGGTAAATTCCTTAAGCCCCATCACACACAAGCTATTATATATTATCATATATTGAAAAACAAACATTAATG
ATGAGGGGAAAATACTGAACAGTCTATAATACACGACTGGAACGATGAGGAGGGCCTTAtAATCTTGAAGAGA
TGCTAAGATGCCGTCGGTATTGgAGGGAAGGTGATTATCATCGATGTGGTTGTGGGTGTCAACCAgGACGTTG
ATGAGGTTCTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTtCAATGCAAAAaACAGAACTGAg
AAAGAATGGGCAAAGTTGATCTCTGATGCTGGCTTCACAGGCTATAACCTTACTCCTGCATTTGGTGTGA

26. *O. americanum* – 9939 (91 bp)

ACTCCTTCAAGCTCAAGCTCATGTATGGAACCATATGTATGCCTTTGCAAACCTCTATGTCCTTAAAATGTGCAAT
TCAATTAGGCATACCCGATATTCTGCACAAACATGACCACCCAATGACACTTTCCCAgTTACTCCAAGCCATCCC
CATCAACAAAGAAAAATCCCAAAGTTTCCAGCGTTTGATGCGTGCACTAGTCAATTCCAATTTCTTCATAGAAG
AAAActCTAATAATCAAGAGGTGTGTTACTGGCTCACCCAGCCTCACGCCTCCTCTTGAAAGGGGCGCCTTT
GACTGTGGCACCCTTGTCCAAGTCGTTTTGGATCCCACTTTCACAAACCCATGGCATTATATGAGTGAATGGT
TTAAACATGAGAACCACGCCACCCAGTTTGAGGCAGCAAATGGATGCACGTTTTGGGAGAAGTTAGCAAATA
AGCCAAGCATGGGTAGATTTTTTATGAAGCTATGAGTTGTGACTCAAGGCTTGTGGCACATGTACTACTAA
GGACTACAAGCATGTGATTGATGGAATAAGAACATTGGTCGATGTTGGGGTGGTAATGGAACGATGGCTAA
AGCTATCGTCGAAGCAGTGGCCACCATGAAATGCACAGTTCTTGACCTACCACATGTTGTGGCTGGGTTGGAA
AGCACCGACAAATTAAGCTATATTGGGGGAGACATGTTCCAGTCTATCCCTTCTGCAGATGCAATTCTTCTCAA
GGTAAATTACTTATAGCCCCATTAAACTTGAATTATTATATATCATTATACATTGAAAAACAAATAATAATGATG
AGTAAAAATACTAAACAGTTTATAATACACGATTGGGACGATGAGGAGGGCCTCAAAATCTTGAAGAGATGTA
AAGATGCAGTTGGCATTGGAGGGAAGGTGATAATCATCGATGTGGTTGTGGGTGTCAACCATGACGTTGATG
AGGTTTTAGAAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCGAAAGAAAGAACTATGAA
TGAATGGGAAAAATTGATTTCTGATGCTGGCTTCACAAGCT

Appendix 3 Full length *O. tenuiflorum* sequence obtained from chapter 4
nucleotide alignment

ATGGCATTGCAAAAAGTAGATATTTCACTATCGACAGAGCAACTTCTTCAAGCTC
AAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATG
TGCAATTCAATTAGGCATACCTGATATTCTCCACAAACATGGCCACCCAATGACA
CTTTCCCAATTACTCCAATCCATCCCCATCAACAAAGCAAAATCCCAATGTCTCC
ATCGTTTGATGCGTGTATTAGTCAATTCCAACCTTCTTCATCGAAGAAAAGAACTC
TAGTAATCAAGAGGTGCATTACTGGCTCACCCAGCCTCGCGCCTCCTCTTGAAA
GGGGCGCCTTTGACAGTGGCACCCCTTGTCOAAGTCATTTTGGATCCCACCTTCA
CAAACCCATGGCACCATATGAGCGAATGGTTTACTCATGAGCACCATGCCACAC
AGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTTAGCAAATGAGCCGA
GCATGGGTAGATTTTTTTGATGAAGCTATGAGTTGTGATGCGAGGCTCGTGGCACA
TGTAATCACTAAGGACTACAAGCATGTGATTGGGGGAATTAGAACGTTGGTTGAT
GTAGGGGGTGGTGACGGAACGATGGCTAAAGCAATCGTTGAAGCAGTGCCCCACC
ATAAAATGCACAGTTATTGACCTCCACATGTCGTGGCTGGCTTGGAAGCACCG
ATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGC
AGTTCTTCTCAAGTCTATAATACACGACTGGAACGATGAGGAGGGCCTTAAAATC
TTGAAGAGATGC_aAAGATGCCGTCGGTATTGGAGGGAAGGTGATTATCATCGATG
TGGTTGTGGGTGTCAACCAGGACGTTGATGAGGTTCTAGAAGATCAGCTCCACTT
CGATATGGCAATGATGTGTTACTTCAATGCAAAAGAA_aAGAACTGAGAAAGAATG
GGCAAAGTTGATCTCTGATGCTGGCTTCACAGGCTATAAGCTTACTCCTGCATTT
GGTGTGAGGTCTCTCATAGAGGCTTATCCTTAA

Appendix 4 Full length *O. gratissimum* sequence obtained from chapter 4
nucleotide alignment

ATGGCATTGCAAAAAGTAGATATTTCACTATCGACAGAGCAACTTCTTCAAGCTC
AAGCTCATGTATGGAACCACATGTACGCCTTTGCAAACCTCAATGTCCTTAAAATG
TGCAATTCAATTAGGCATTCCCGATATTCTCCACAAACATGGCCACCCAATGACA
CTTTCCCAATTACTCCAATCCATCCCCATTAACAAAGCAAAATCCCAATGTCTCC
ATCGTTTGATGCGTGTATTAGTCAATTCCAATTTCTTCATCGAGGAAAACAACCTCT
AATAATCAAGAGGTGCGTTACTGGCTGACCCAGCCTCACGCCTCCTCTTGAAAG
GGGCGCCTCTGACTGTGGCACCCCTTGTCOAAGTCATTTTGGATCCCACCTTTCAC
AAACCCATGGCACCATATGAGCGAATGGCTTACACA_cGAGCACACACACACACA
GTTTGAGGCAGCAAATGGATGCATGTTTTGGGAGAAGTTAGCAAATGAGCCGAG
CATGAGTAGATTTTTTTGATGAAGCTATGAGTTGTGACTCGAGGCTCGTGGCACAT
GTACTACCAAGGACTACAAGCATGTGATTGAGGGAATCAAACATTGGTTGAT
GTAGGGGGTGGTAATGGAACGATGGCTA_gAGCTATCGTTGAAGCAGTGCCCCACC
ATGAAATGCACAGTTATTGACCTCCACATGTCGTGGCTGGCTTGGAAGCACTG
ATAGCTTGAGCTATATTGGGGGAGACATGTTCCGGTCTATCCCTTCTGCCGATGC
AGTTCTTCTCAAGTCTATAATACACGACAGGAAAGATAAGGAAGGCCTCAAAAT
CTTGAATAGATACAAAGATGCAGTCGGTATTGGAGGTAAGGTGATTATTATCGAC
GTGGTTGTGGGTGTCAATCATGATGTTGATAAGGTTCTAGAAGATCAGCTCCACT
TCGATATGGCAATGATGTGTTACTTCAATGCGAAAGAAAGAACTGAGAAAGAAT
GGGCAAAGTTGATCTCTGATGCTGGCTTCACAAGCTATAAACTTACTCCTGCATT
TGGTGTGAGGTCTCTCATAGAGGCTTATCCTTAA

Appendix 5 G-blocks

1. *O. basilicum* – EOMT

AAAAAA**GCGGCCGC**ATGGCATTGCAAAAAGTAGATATTTTCATTGTGCGACAGAGC
AACTTCTTCAAGCTCAAGTTCATGTATGGAACCATATGTATGCCTTTGCAAATTC
AATGTCCCTAAAGTGTGCAATTCAATTAGGCATACCCGATATTCTCCACAAACAT
GGCCGCCCAATGACACTTTCCCAGTTACTCCAATCCATCCCCATCAACAAAGAAA
AAACCCAATGTTTCCAACGTTTGATGCGTGCCTAGTCAATTCAAATTTTTTCATC
GAAGAAAACAACCTCTAATAATCAAGAGGTGTGCTACTGGCTCACCCAGCGTCA
TGCCTCCTCTTGAAGGAGGCGCCCCCTAACTGTGACACCCCTAGTCCAAGTCGTTT
TGGATCCCCTTTTCAAAACCCATGGCACCATATGAGTGAATGGTTTACACATGA
GAAACATGCCACACAGTTTGAGGCAGCAAACGGATGCACATTTTGGGAGAAGTT
AGCAAATGAGCCAAGCAAGGGTAGATTTTTTTGATGAAGCTATGAGTTGTGACTC
GAGGCTCATAGCACATGTATTCACCAAGGACTACAAGCATGTGATTGAGGGAAT
CAGAACATTGGTTGATGTTGGTGGTGAATGGAACGATGGCTAAAGCTATCGTT
GAAGCAATGCCACCATTAATGCACAGTTATTGACCTCCCACATGTTGTGGCTG
GCTTGGAAGCACCGATAACTTAACTATATTGGAGGAGACATGTTCCAGTCTAT
CCCTTCTGCAGATGCAATTCTTCTAAAGTCTATAATACATGATTGGGACGATGTG
GAGGGCCTCAAAATCTTGAAGAAATGCAAAGATGCGGTCGTTATGGGAGGGAAG
GTGATTATCATCGACGTGGTTGTGGGTGTCAACCACGATATTGATGAGGTTTTAG
AAGATCAGCTTCACTTTGATATGGCAATGATGTGTTACTTCAATGCGAAAGAAAG
AACTATGAGTGAATGGGAAAAGTTGATTTATGATGCTGGCTTCAAAGCTATAA
GCTTACTCCTGCATTTGGTGTAAGGTCTCTCATAGAGGCTTATCCTTAA**GAATTC**A
AAAAA

2. *O. basilicum* – CVOMT

AAAAAA**GCGGCCGC**ATGGCATTGCAAAATATGGATATTTTCATTGTGCGACTGAGC
AACTCCTTCAAGCTCAAGCTCATGTATGGAACCATATGTATGCCTTTGCAAATTC
TATGTCCCTTAAATGTGCAATTCAATTAGGCATACCCGATATTCTGCACAAACAT
GACCACCCAATGACACTTTCCCAATTACTCAAGGCCATCCCCATCAACAAAGAA
AAATCCCAAAGTTTCCAGCGTTTGATGCGTGCCTAGTCAACTCCAATTTCTTCA
TAGAAGAAAACCTCTAATAATCAAGAGGTGTGTTACTGGCTCACCCAGCCTCAC
GCCTCCTCTTGAAGGGGGCGCCTTTGACTGTGGCACCCCTTGTTCAAGTGGTTTT
GGATCCCCTTTTCAAAACCCATGGCATTATATGAGTGAATGGTTTAAACATGAG
AACCACGCCACCCAGTTTGAGGCAGCAAATGGATGCACGTTTTGGGAGAAGTTA
GCAAATAAGCCCAGCATGGGTAGATTTTTTTGATGAAGCTATGAGTTGTGACTCAA
GGCTTGTGGCACATGTACTACTAAGGACTACAAGCATGTGATTGATGGAATAA
GAACATTGGTCGATGTTGGGGGTGGTAATGGAACGATGGCTAAAGCTATCGTCG
AAGCAGTGCCACCATGAAATGCACTGTTCTTGACCTACCACATGTTGTGGCTGG
GTTGGAAAGCACCGACAAATTAAGCTATATTGGGGGAGACATGTTCCAGTCTAT
CCCTTCTGCAGATGCAATTCTTCTCAAGTTTATAATACACGATTGGGACGATGAG
GAGGGCCTCAAAATCTTGAAGAGATGTAAAGATGCAGTTGGCATTGGAGGGAAG
GTGATAATCATCGATGTGGTTGTGGGTGTCAACCATGACGTTGATGAGGTTTTAG
AAGATCAACTCCACTTCGATATGGCAATGATGTCTTACTTCAATGCGAAAGAAAG
AACTATGAATGAATGGGAAAAATTGATTTCTGCTGCTGGCTTACAAGCTATAAG
CTTACTCCAGCATTTGGTGAGGTCTCTCATAGAGGCTTATCCTTAA**GAATTC**A
AAAAA

3. *O. tenuiflorum*

AAAAAA**GCGGCCGC**ATGGCATTGCAAAAAGTAGATATTTCACTATCGACAGAGC
AACTTCTTCAAGCTCAAGCTCATGTGTGGAATCACATGTACGCCTTTGCAAACATC
AATGTCCTTAAAATGTGCAATTCAATTAGGCATACCTGATATTCTCCACAAACAT
GGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCCCCATCAACAAAGCAA
AATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAACCTTCTTCATC
GAAGAAAAGAAGCTCTAGTAATCAAGAGGTGCATTACTGGCTCACCCCAGCCTCG
CGCCTCCTCTTGAAAGGGGCGCCTTTGACAGTGGCACCCCTTGTTCCAAGTCATTT
TGGATCCCACCTTTCACAAACCCATGGCACCATATGAGCGAATGGTTTACTCATGA
GCACCATGCCACACAGTTTGAGGCGGCAAATGGATGCATGTTTTGGGAAAAGTT
AGCAAATGAGCCGAGCATGGGTAGATTTTTTGATGAAGCTATGAGTTGTGATGCG
AGGCTCGTGGCACATGTACTCACTAAGGACTACAAGCATGTGATTGGGGGAATT
AGAACGTTGGTTGATGTAGGGGGTGGTGACGGAACGATGGCTAAAGCAATCGTT
GAAGCAGTGCCCAACATAAAATGCACAGTTATTGACCTCCCACATGTCGTGGCTG
GCTTGGAAGCACCAGATAGCTTGAGCTATATTGGAGGAGACATGTTCCGGTCTAT
CCCTTCTGCCGATGCAGTTCTTCTCAAGTCTATAATACACGACTGGAACGATGAG
GAGGGCCTTAAAATCTTGAAGAGATGC_aAAGATGCCGTCGGTATTGGAGGGAAG
GTGATTATCATCGATGTGGTTGTGGGTGTCAACCAGGACGTTGATGAGGTTCTAG
AAGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCAAAAGAA_aAG
AACTGAGAAAGAATGGGCAAAGTTGATCTCTGATGCTGGCTTCACAGGCTATAA
GCTTACTCCTGCATTTGGTGTGAGGTCTCTCATAGAGGCTTATCCTTAA**GAATTCA**
AAAAA

4. *O. gratissimum*

AAAAAA**GCGGCCGC**ATGGCATTGCAAAAAGTAGATATTTCACTATCGACAGAGC
AACTTCTTCAAGCTCAAGCTCATGTATGGAACCATGTACGCCTTTGCAAACATC
AATGTCCTTAAAATGTGCAATTCAATTAGGCATTCCCGATATTCTCCACAAACAT
GGCCACCCAATGACACTTTCCCAATTACTCCAATCCATCCCCATTAACAAAGCAA
AATCCCAATGTCTCCATCGTTTGATGCGTGTATTAGTCAATTCCAATTTCTTCATC
GAGGAAAACAACTCTAATAATCAAGAGGTGCGTTACTGGCTGACCCCAGCCTCA
CGCCTCCTCTTGAAAGGGGGCGCCTCTGACTGTGGCACCCCTTGTTCCAAGTCATTT
TGGATCCCACCTTTCACAAACCCATGGCACCATATGAGCGAATGGCTTACACA_cGA
GCACCACACCACACAGTTTGAGGCAGCAAATGGATGCATGTTTTGGGAGAAAGTT
AGCAAATGAGCCGAGCATGAGTAGATTTTTTGATGAAGCTATGAGTTGTGACTCG
AGGCTCGTGGCACATGTACTACCAAGGACTACAAGCATGTGATTGAGGGAATC
AAAACATTGGTTGATGTAGGGGGTGGTAATGGAACGATGGCTA_gAGCTATCGTTG
AAGCAGTGCCCAACCATGAAATGCACAGTTATTGACCTCCCACATGTCGTGGCTGG
CTTGGAAGCACTGATAGCTTGAGCTATATTGGGGGAGACATGTTCCGGTCTATC
CCTTCTGCCGATGCAGTTCTTCTCAAGTCTATAATACACGACAGGAAAGATAAGG
AAGGCCTCAAAATCTTGAATAGATACAAAGATGCAGTCGGTATTGGAGGTAAGG
TGATTATTATCGACGTGGTTGTGGGTGTCAATCATGATGTTGATAAGGTTCTAGA
AGATCAGCTCCACTTCGATATGGCAATGATGTGTTACTTCAATGCGAAAGAAAG
AACTGAGAAAGAATGGGCAAAGTTGATCTCTGATGCTGGCTTCACAAGCTATAA
ACTTACTCCTGCATTTGGTGTGAGGTCTCTCATAGAGGCTTATCCTTAA**GAATTCA**
AAAAA

Appendix 6 *Ocimum* (EOMT) Sequence Accession Numbers from GenBank

1. *Ocimum basilicum* Sequence 493 (B16) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442468)
2. *Ocimum basilicum* Sequence 745 (G34) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442469)
3. *Ocimum citriodorum* Sequence 674 (G29) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442470)
4. *Ocimum americanum* Sequence 9939 (C85) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442471)
5. *Ocimum kilimandscharicum* Sequence 676 (G31) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442472)
6. *Ocimum selloi* Sequence 1583 (C82) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442473)
7. *Ocimum gratissimum* Sequence 744 (G30) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442474)
8. *Ocimum gratissimum* Sequence 769 (G32) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442475)
9. *Ocimum gratissimum* Sequence 1138 (C81) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442476)
10. *Ocimum tenuiflorum* Sequence 770 (G37) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442477)
11. *Ocimum tenuiflorum* Sequence 285 (14092902) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442478)
12. *Ocimum tenuiflorum* Sequence 286 (14092903) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442479)
13. *Ocimum tenuiflorum* Sequence 292 (14092909) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442480)
14. *Ocimum tenuiflorum* Sequence 618 (G26) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442481)
15. *Ocimum tenuiflorum* Sequence 619 (G27) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442482)
16. *Ocimum tenuiflorum* Sequence 746 (G36) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442483)
17. *Ocimum tenuiflorum* Sequence 771 (G38) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442484)
18. *Ocimum tenuiflorum* Sequence 844 (G39) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442485)
19. *Ocimum tenuiflorum* Sequence 865 (V86) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442486)
20. *Ocimum tenuiflorum* Sequence 866 (V87) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442487)
21. *Ocimum tenuiflorum* Sequence 867 (V88) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442488)
22. *Ocimum tenuiflorum* Sequence 871 (V92) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442489)
23. *Ocimum tenuiflorum* Sequence 872 (V93) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442490)

24. *Ocimum tenuiflorum* Sequence 873 (V94) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442491)
25. *Ocimum tenuiflorum* Sequence 877 (V98) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442492)
26. *Ocimum tenuiflorum* Sequence 878 (V99) eugenol O-methyltransferase gene, partial sequence (Accession No. - MK442493)